

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 14:54:57 ; Search time 28.5224 Seconds  
(without alignments)  
69.343 Million cell updates/sec

Title: US-09-641-801-1  
Perfect score: 42  
Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	42	100.0	7	4	AAB72500	Aab72500 Colostrin
2	42	100.0	7	4	AAB59324	Aab59324 Ewe colos
3	42	100.0	7	4	AAB72246	Aab72246 Colostrin
4	42	100.0	7	4	AAB72532	Aab72532 Colostrin
5	42	100.0	7	5	AAO14577	Aao14577 Neural ce
6	42	100.0	7	5	AAM51036	Aam51036 Colostrin
7	42	100.0	7	5	AAE20228	Aae20228 Colostrin
8	42	100.0	8	4	AAB59354	Aab59354 Ewe colos
9	42	100.0	316	4	ABB67510	Abb67510 Drosophil

10	39	92.9	95	4	AAU45697	Aau45697	Propionib
11	39	92.9	95	6	ABM42216	Abm42216	Propionib
12	39	92.9	132	4	AAM17529	Aam17529	Peptide #
13	39	92.9	132	4	ABB36552	Abb36552	Peptide #
14	39	92.9	132	4	AAM30051	Aam30051	Peptide #
15	39	92.9	132	4	ABB31347	Abb31347	Peptide #
16	39	92.9	132	4	ABB21889	Abb21889	Protein #
17	39	92.9	132	4	AAM69716	Aam69716	Human bon
18	39	92.9	132	4	AAM57317	Aam57317	Human bra
19	39	92.9	132	4	ABG51400	Abg51400	Human liv
20	39	92.9	132	4	AAM05201	Aam05201	Peptide #
21	39	92.9	132	5	ABG39337	Abg39337	Human pep
22	38	90.5	515	5	ABB90945	Abb90945	Herbicida
23	37	88.1	38	4	AAE10123	Aae10123	Human ion
24	37	88.1	38	5	AAU83499	Aau83499	Novel hum
25	37	88.1	38	6	ABU97384	Abu97384	Amino aci
26	37	88.1	38	7	ADE29306	Ade29306	Novel hum
27	37	88.1	70	5	ABP10528	Abp10528	Human ORF
28	37	88.1	73	3	AAB41170	Aab41170	Human ORF
29	37	88.1	73	5	ABP01103	Abp01103	Human ORF
30	37	88.1	95	4	ABG21168	Abg21168	Novel hum
31	37	88.1	99	3	AAG23978	Aag23978	Arabidops
32	37	88.1	186	7	ADB37401	Adb37401	Stem cell
33	37	88.1	278	2	AAR79095	Aar79095	Rat Fas l
34	37	88.1	278	2	AAW98069	Aaw98069	Rat Fas l
35	37	88.1	278	2	AAW95040	Aaw95040	Rat FasL
36	37	88.1	278	5	AAO19029	Aao19029	Human Fas
37	37	88.1	278	7	ADD69944	Add69944	Rat Fas l
38	37	88.1	302	3	AAV50933	Aay50933	Human fet
39	37	88.1	320	5	AAU78461	Aau78461	Mouse bet
40	37	88.1	332	4	AAE10122	Aae10122	Human ion
41	37	88.1	332	6	ABU97382	Abu97382	Amino aci
42	37	88.1	332	7	ADE29303	Ade29303	Novel hum
43	37	88.1	354	5	ABP41838	Abp41838	Human ova
44	37	88.1	358	7	ADB64222	Adb64222	Human pro
45	37	88.1	393	4	AAB83413	Aab83413	Human ion

# ALIGNMENTS

RESULT 1

AAB72500

ID AAB72500 standard; peptide; 7 AA.

XX

AC AAB72500;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #1.

XX

KW Dermatological; oxidative stress regulator; colostrinin.

XX

OS Unidentified.

XX

PN WO200112650-A2.

XX

PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022665.  
 XX  
 PR 17-AUG-1999; 99US-0149310P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I;  
 XX  
 DR WPI; 2001-218342/22.  
 XX  
 PT Modulating oxidative stress level in a cell, involves contacting the cell  
 PT with an oxidative stress regulator selected from colostrinin, its  
 PT constituent peptide, analog or their combinations.  
 XX  
 PS Claim 6; Page 25; 48pp; English.  
 XX  
 CC The present invention relates to a method for modulating the oxidative  
 CC stress level in a cell or a patient, comprising contacting the cell with,  
 CC or administering to the patient, an oxidative stress regulator selected  
 CC from colostrinin, or its constituent peptide (e.g. the present peptide),  
 CC to change the level of an oxidising species in the cell. The method can  
 CC be used to treat oxidative damage to skin, by decreasing or preventing an  
 CC increase in the level of damage to a biomolecule of the patient  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7

# RESULT 2

AAB59324

ID AAB59324 standard; peptide; 7 AA.

XX

AC AAB59324;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment B-9.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;

KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX  
 PR 02-JUN-1999; 99GB-00012852.  
 XX  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Georgiades JA;  
 XX  
 DR WPI; 2001-071058/08.  
 XX  
 PT Peptides having an N-terminal amino acid sequence isolated from  
 PT colostrinin for treating e.g. disorders of the central nervous system and  
 PT immune system, viral and bacterial infections, and diseases characterized  
 PT by amyloid plaques.  
 XX  
 PS Claim 7; Page 27; 63pp; English.  
 XX  
 CC The present invention provides the sequences of a number of peptides  
 CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
 CC fragment of colostrum. These peptides can be used in the treatment of  
 CC central nervous system disorders such as senile dementia, Parkinson's  
 CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
 CC disorders such as bacterial and viral infections, to improve the  
 CC development of a child's immune system, as a dietary supplement, and to  
 CC promote the dissolution of beta-amyloid plaques  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7

# RESULT 3

AAB72246

ID AAB72246 standard; peptide; 7 AA.

XX

AC AAB72246;

XX

DT 14-MAY-2001 (first entry)

XX

DE Colostrinin derived cytokine inducing peptide SEQ ID 1.

XX

KW Colostrinin; immune response; cytokine; blood cell proliferation;

KW central nervous system disorder; neurological disorder; mental disorder;

KW dementia; neurodegenerative disease; Alzheimer's disease; psychosis;

KW neurosis; infection.

XX

OS Synthetic.

XX

PN WO200111937-A2.

XX

PD 22-FEB-2001.

XX

PF 17-AUG-2000; 2000WO-US022818.  
 XX  
 PR 17-AUG-1999; 99US-0149311P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2001-202804/20.  
 XX  
 PT Inducing a cytokine and modulating an immune response, useful for  
 PT treating central nervous system diseases and bacterial and viral  
 PT infections, comprises administering colostrinin as an immunological  
 PT regulator.  
 XX  
 PS Claim 1; Page 34; 50pp; English.  
 XX  
 CC Sequences AAB72246 - AAB72275 represent peptides derived from clostrinin,  
 CC a proline rich polypeptide aggregate contained in colostrum. The peptides  
 CC have immune response modulatory activity, and are capable of inducing  
 CC cytokines. Colostrinin and its derived peptides are useful for inducing  
 CC cytokine production, for modulating an immunological response and for  
 CC inducing blood cell proliferation. The peptides are useful in the  
 CC treatment of disorders of the central nervous system, neurological  
 CC disorders, mental disorders, dementia, neurodegenerative diseases,  
 CC Alzheimer's disease, motor neurone disease, psychosis, neurosis, chronic  
 CC disorders of the immune system, bacterial and viral infections and  
 CC acquired immunological deficiencies  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7

#### RESULT 4

AAB72532

ID AAB72532 standard; peptide; 7 AA.

XX

AC AAB72532;

XX

DT 09-MAY-2001 (first entry)

XX

DE Colostrinin peptide #1.

XX

KW Neuroprotective; neural cell differentiation regulator; colostrinin;  
 KW colostrum.

XX

OS Unidentified.

XX

PN WO200112651-A2.

XX  
 PD 22-FEB-2001.  
 XX  
 PF 17-AUG-2000; 2000WO-US022774.  
 XX  
 PR 17-AUG-1999; 99US-0149633P.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I;  
 XX  
 DR WPI; 2001-226545/23.  
 XX  
 PT Use of colostrinin, its constituent peptide or analog as a neural cell  
 PT regulator, for promoting neural cell differentiation and treating damaged  
 PT neural cells in a patient.  
 XX  
 PS Claim 6; Page 21; 35pp; English.  
 XX  
 CC The present invention relates to a method for promoting neural cell  
 CC differentiation and treating damaged neural cells, using colostrinin and  
 CC colostrinin constituent peptides (e.g. the present peptide) as a neural  
 CC cell regulator. Colostrinin is a polypeptide complex found in colostrum  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7

# RESULT 5

AAO14577

ID AAO14577 standard; peptide; 7 AA.

XX

AC AAO14577;

XX

DT 27-MAY-2002 (first entry)

XX

DE Neural cell regulatory colostrinin peptide 1.

XX

KW Neural cell differentiation; neural cell regulator; colostrinin peptide;

KW neural cell formation; proline-rich polypeptide aggregate; colostrum;

KW neural cell treatment.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 7

FT /note= "Optional C-terminal amide"

XX

PN WO200213851-A1.

XX

PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022777.  
 XX  
 PR 17-AUG-2000; 2000WO-US022777.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Boldogh I, Stanton JG, Hughes TK;  
 XX  
 DR WPI; 2002-269152/31.  
 XX  
 PT Promoting cell differentiation in a patient involves use of blood cell  
 PT regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 7; Page 21; 37pp; English.  
 XX  
 CC The invention comprises a method for promoting cell differentiation (e.g.  
 CC neural cell differentiation). The method involves contacting cells with a  
 CC neural cell regulator (i.e. a colostrinin peptide) in order to change the  
 CC cells in morphology to form neural cells. Colostrinin is a proline-rich  
 CC polypeptide aggregate that is present in colostrum. The method of the  
 CC invention is useful for promoting the differentiation of cells and for  
 CC treating damaged neural cells in a patient. The present amino acid  
 CC sequence represents a specifically claimed colostrinin peptide used in  
 CC the method of the invention  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7

# RESULT 6

AAM51036

ID AAM51036 standard; peptide; 7 AA.  
 XX  
 AC AAM51036;  
 XX  
 DT 30-MAY-2002 (first entry)  
 XX  
 DE Colostrinin constituent peptide.  
 XX  
 KW Colostrinin; colostrum; immunomodulator; cardiovascular;  
 KW blood cell regulator; cytokine inducer; beta-casein; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 7  
 FT /note= "optional C-terminal amidation"

XX  
 PN WO200213849-A1.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 17-AUG-2000; 2000WO-US022775.  
 XX  
 PR 17-AUG-2000; 2000WO-US022775.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 PA (REGE-) REGEN THERAPEUTICS PLC.  
 XX  
 PI Stanton GJ, Hughes TK, Boldogh I, Georgiades J;  
 XX  
 DR WPI; 2002-269150/31.  
 XX  
 PT Modulation of blood cell proliferation in a patient involves use of blood  
 PT cell regulator selected from colostrinin, its constituent peptide and/or  
 PT analog.  
 XX  
 PS Claim 1; Page 34; 54pp; English.  
 XX  
 CC The present sequence is that of a colostrinin constituent peptide that is  
 CC preferred for use as an immunological regulator and as a blood cell  
 CC regulator in claimed methods of the invention. It is classified as having  
 CC a beta-casein homologue precursor. Methods are claimed for: inducing a  
 CC cytokine in a cell by contact with an immunological regulator, where the  
 CC cell is present in a cell culture, a tissue, an organ or an organism, and  
 CC the cell is mammalian, including human; modulating an immune response in  
 CC a cell by contact with the immunological regulator under conditions  
 CC effective to induce a cytokine; modulating an immune response in a  
 CC patient by administering an immunological regulator under conditions  
 CC effective to induce a cytokine, where the immunological regulator is  
 CC administered topically or as part of a dietary supplement, and where the  
 CC immune response is specific or non specific, an interferon response or an  
 CC antibody response; modulating blood cell proliferation by contacting  
 CC blood cells with a blood cell regulator, where the blood cells are  
 CC present in a cell culture or an organism, are mammalian or human, and  
 CC where the blood cells are increased in number or differentiated; and a  
 CC method for modulating blood cell proliferation in a patent. A claimed  
 CC cytokine-inducing composition comprises a pharmaceutical carrier and an  
 CC active agent such as the present peptide. Cytokines induced by this  
 CC peptide in human leucocyte cultures include interferon-gamma, tumour  
 CC necrosis factor-alpha, interleukin-4, interleukin-6, interleukin-10 and  
 CC interleukin-12. It was one of the best overall inducers in almost all  
 CC cytokine and blood cell proliferation experiments conducted  
 XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQPPPLP 7  
 |||||  
 Db 1 MQPPPLP 7



RESULT 7

AAE20228

ID AAE20228 standard; peptide; 7 AA.

XX

AC AAE20228;

XX

DT 18-JUN-2002 (first entry)

XX

DE Colostrinin constituent peptide #1.

XX

KW Blood cell regulator; colostrinin; constituent peptide; oxidative stress;  
KW therapy; oxidative damage; skin; aging; wound healing; cell replacement;  
KW tissue; organ; cosmetic procedure; repair; regeneration; preservation;  
KW transplantation; implantation; dermatological; vulnerary.

XX

OS Unidentified.

XX

FH Key Location/Qualifiers

FT Modified-site 7

FT /note= "Optionally C-terminal amide"

XX

PN WO200213850-A1.

XX

PD 21-FEB-2002.

XX

PF 17-AUG-2000; 2000WO-US022776.

XX

PR 17-AUG-2000; 2000WO-US022776.

XX

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Stanton GJ, Hughes TK, Boldogh I;

XX

DR WPI; 2002-269151/31.

XX

PT Composition useful for the modulation of blood cell proliferation in a  
PT patient comprises a blood cell regulator selected from colostrinin, its  
PT constituent peptide and/or analog.

XX

PS Claim 6; Page 25; 51pp; English.

XX

CC The invention relates to a composition which comprises a blood cell  
CC regulator selected from colostrinin, its constituent peptide and/or  
CC analogue. The invention is used for modulating the oxidative stress level  
CC in a cell e.g. mammalian or human cell present in a cell culture, tissue,  
CC organ, or organism; or for treating oxidative damage to the skin of a  
CC patient e.g. animal or human; to modulate oxidative stress during/ after  
CC a premature birth or normal birth, preventing/delaying aging in a  
CC patient, enhancing wound healing, and the reduction of side effects of  
CC cosmetic procedures. The method changes the level of an oxidising species  
CC in the cell, such as decreases or prevents increase in the level of  
CC damage to a biomolecule of the patient selected from DNA, protein and/or  
CC lipid, compared to the same conditions when the oxidative stress  
CC regulator is not present. The modulation of oxidative stress results in  
CC enhanced repair, regeneration, and replacement of cells, tissues and  
CC organs (e.g. kidney, liver, pancreas, skin, and the other internal and

CC external organs), as well as enhanced preservation of such organs for  
CC transplantation, implantation, or scientific research. The present  
CC sequence is a colostrinin constituent peptide  
XX  
SQ Sequence 7 AA;

Query Match 100.0%; Score 42; DB 5; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | | | |  
Db 1 MQPPPLP 7

RESULT 8

AAB59354

ID AAB59354 standard; peptide; 8 AA.

XX

AC AAB59354;

XX

DT 21-MAR-2001 (first entry)

XX

DE Ewe colostrinin peptide fragment derived sequence #14.

XX

KW Sheep; colostrinin; proline rich polypeptide; colostrum; immune disorder;  
KW central nervous system disorder; dietary supplement; beta-amyloid plaque.

XX

OS Ovis sp.

XX

PN WO200075173-A2.

XX

PD 14-DEC-2000.

XX

PF 02-JUN-2000; 2000WO-GB002128.

XX

PR 02-JUN-1999; 99GB-00012852.

XX

PA (REGE-) REGEN THERAPEUTICS PLC.

XX

PI Georgiades JA;

XX

DR WPI; 2001-071058/08.

XX

PT Peptides having an N-terminal amino acid sequence isolated from  
PT colostrinin for treating e.g. disorders of the central nervous system and  
PT immune system, viral and bacterial infections, and diseases characterized  
PT by amyloid plaques.

XX

PS Claim 8; Page 27; 63pp; English.

XX

CC The present invention provides the sequences of a number of peptides  
CC found in ewe's colostrinin. Colostrinin is the proline-rich polypeptide  
CC fragment of colostrum. These peptides can be used in the treatment of  
CC central nervous system disorders such as senile dementia, Parkinson's  
CC disease, Alzheimer's disease, psychosis and neurosis, immune system  
CC disorders such as bacterial and viral infections, to improve the

CC development of a child's immune system, as a dietary supplement, and to  
CC promote the dissolution of beta-amyloid plaques  
XX  
SQ Sequence 8 AA;

Query Match 100.0%; Score 42; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | |  
Db 2 MQPPPLP 8

RESULT 9

ABB67510

ID ABB67510 standard; protein; 316 AA.

XX

AC ABB67510;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 29322.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11613.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

XX

PS Disclosure; SEQ ID NO 29322; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 316 AA;

Query Match 100.0%; Score 42; DB 4; Length 316;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | | | |  
Db 249 MQPPPLP 255

RESULT 10  
AAU45697

ID AAU45697 standard; protein; 95 AA.

XX

AC AAU45697;

XX

DT 27-FEB-2002 (first entry)

XX

DE Propionibacterium acnes immunogenic protein #6593.

XX

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

XX

OS Propionibacterium acnes.

XX

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US012865.

XX

PR 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.

PR 07-JUL-2000; 2000US-0216747P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59526.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

XX

PS Claim 6; SEQ ID NO 6892; 1069pp; English.

XX

CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC *P. acnes* is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 95 AA;

Query Match 92.9%; Score 39; DB 4; Length 95;  
Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|:|||||  
Db 88 MEPPPLP 94

RESULT 11

ABM42216

ID ABM42216 standard; protein; 95 AA.

XX

AC ABM42216;

XX

DT 20-OCT-2003 (first entry)

XX

DE *Propionibacterium acnes* immunogenic polypeptide #6892.

XX

KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;

KW immunostimulant; immune response; vaccine; immunogenic.

XX

OS *Propionibacterium acnes*.

XX

PN WO2003033515-A1.

XX

PD 24-APR-2003.

XX

PF 11-OCT-2002; 2002WO-US032727.

XX

PR 15-OCT-2001; 2001US-00978825.

XX

PA (CORI-) CORIXA CORP.

XX

PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieve-Douglass J;  
 XX  
 DR WPI; 2003-381789/36.  
 DR N-PSDB; ACF64455.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Claim 6; SEQ ID NO 6892; 1481pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a specifically claimed P. acnes polypeptide which is  
 CC thought to contain an immunogenic region. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 95 AA;

Query Match 92.9%; Score 39; DB 6; Length 95;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQPPPLP 7  
 |:|||||  
 Db 88 MEPPPLP 94

RESULT 12  
 AAM17529  
 ID AAM17529 standard; protein; 132 AA.  
 XX  
 AC AAM17529;

XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Peptide #3963 encoded by probe for measuring cervical gene expression.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157278-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000670.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human cervical epithelial cells.  
 XX  
 PS Claim 27; SEQ ID NO 22355; 487pp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 132 AA;

Query Match 92.9%; Score 39; DB 4; Length 132;  
 Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 116 LQPPPLP 122

RESULT 13

ABB36552

ID ABB36552 standard; peptide; 132 AA.

XX

AC ABB36552;

XX

DT 04-FEB-2002 (first entry)

XX

DE Peptide #4058 encoded by human foetal liver single exon probe.

XX

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX

OS Homo sapiens.

XX

PN WO200157277-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000669.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human fetal liver.

XX

PS Claim 27; SEQ ID NO 29187; 639pp + Sequence Listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human fetal liver. The  
CC present sequence is a peptide encoded by a single exon nucleic acid probe  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 132 AA;

Query Match 92.9%; Score 39; DB 4; Length 132;

Best Local Similarity 85.7%; Pred. No. 3.1e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7

:|||||

Db 116 LQPPPLP 122



RESULT 14

AAM30051

ID AAM30051 standard; protein; 132 AA.

XX

AC AAM30051;

XX

DT 17-OCT-2001 (first entry)

XX

DE Peptide #4088 encoded by probe for measuring placental gene expression.

XX

KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.

XX

OS Homo sapiens.

XX

PN WO200157272-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000663.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488897/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.

XX

PS Claim 27; SEQ ID NO 30320; 654pp; English.

XX

CC The present invention relates to single exon nucleic acid probes (SENP:  
CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders

XX

SQ Sequence 132 AA;

Query Match 92.9%; Score 39; DB 4; Length 132;

Best Local Similarity 85.7%; Pred. No. 3.1e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7

:|||||

## RESULT 15

ABB31347

ID ABB31347 standard; peptide; 132 AA.

XX

AC ABB31347;

XX

DT 01-FEB-2002 (first entry)

XX

DE Peptide #3998 encoded by breast cell single exon nucleic acid probe.

XX

KW Human; microarray; single exon probe; gene expression; breast; disease;  
cancer.

XX

OS Homo sapiens.

XX

PN WO200157271-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000662.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-496933/54.

XX

PT New spatially-addressable set of single exon nucleic acid probes, useful  
PT for measuring gene expression in sample derived from human breast,  
PT comprises number of single exon nucleic acid probes.

XX

PS Claim 27; SEQ ID NO 14315; 327pp + Sequence Listing; English.

XX

CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting the  
CC probes with a collection of detectably labelled nucleic acids derived  
CC from mRNA of human breast, and then measuring the label bound to each  
CC probe of the microarray. The probes are useful for verifying the  
CC expression of regions of genomic DNA predicted to encode proteins. They  
CC are useful for gene discovery, and for determining predisposition and/or  
CC prognosing breast disease. Gene expression analysis is useful for  
CC assessing the toxicity of chemical agents on cells. The microarray of  
CC this invention presents a far greater diversity of probes for measuring  
CC gene expression, with far less bias than expressed sequence tag  
CC microarrays. The method is suitable for rapid production of functional

CC information from genomic sequence. The present sequence is a peptide  
CC encoded by a single exon nucleic acid probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 132 AA;

Query Match 92.9%; Score 39; DB 4; Length 132;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 116 LQPPPLP 122

Search completed: August 24, 2004, 15:42:03  
Job time : 33.5224 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:33:13 ; Search time 7.6791 Seconds  
 (without alignments)  
 47.060 Million cell updates/sec

Title: US-09-641-801-1  
 Perfect score: 42  
 Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	42	100.0	7	4	US-09-641-803-1	Sequence 1, Appli
2	37	88.1	278	4	US-08-339-214-16	Sequence 16, Appl
3	37	88.1	278	4	US-08-339-214-26	Sequence 26, Appl
4	37	88.1	278	4	US-09-131-237C-6	Sequence 6, Appli
5	37	88.1	479	4	US-09-252-991A-31536	Sequence 31536, A
6	37	88.1	642	4	US-09-252-991A-32104	Sequence 32104, A
7	37	88.1	667	2	US-08-718-661-2	Sequence 2, Appli
8	37	88.1	721	4	US-09-533-029-78	Sequence 78, Appl
9	36	85.7	349	1	US-08-118-270-71	Sequence 71, Appl
10	36	85.7	349	5	PCT-US93-08528-71	Sequence 71, Appl
11	35	83.3	23	1	US-08-268-251-34	Sequence 34, Appl

12	35	83.3	23	5	PCT-US93-01112-34	Sequence 34, Appl
13	35	83.3	447	4	US-09-388-349-9	Sequence 9, Appli
14	34	81.0	76	4	US-09-096-731A-14	Sequence 14, Appl
15	34	81.0	158	4	US-09-252-991A-24956	Sequence 24956, A
16	34	81.0	194	3	US-08-822-264-4	Sequence 4, Appli
17	34	81.0	194	4	US-09-565-808-21	Sequence 21, Appl
18	34	81.0	195	2	US-08-960-022-6	Sequence 6, Appli
19	34	81.0	195	4	US-09-565-808-2	Sequence 2, Appli
20	34	81.0	223	3	US-08-822-264-3	Sequence 3, Appli
21	34	81.0	223	4	US-09-565-808-22	Sequence 22, Appl
22	34	81.0	235	4	US-09-370-950C-2	Sequence 2, Appli
23	34	81.0	236	4	US-09-370-950C-3	Sequence 3, Appli
24	34	81.0	256	4	US-09-252-991A-20981	Sequence 20981, A
25	34	81.0	258	4	US-09-096-731A-18	Sequence 18, Appl
26	34	81.0	304	4	US-09-096-731A-10	Sequence 10, Appl
27	34	81.0	309	4	US-09-489-039A-11604	Sequence 11604, A
28	34	81.0	376	4	US-09-056-556-202	Sequence 202, App
29	34	81.0	376	4	US-09-072-596-197	Sequence 197, App
30	34	81.0	376	4	US-09-072-967-202	Sequence 202, App
31	34	81.0	409	4	US-09-029-755C-2	Sequence 2, Appli
32	34	81.0	608	3	US-09-413-814-92	Sequence 92, Appl
33	34	81.0	690	4	US-09-252-991A-16715	Sequence 16715, A
34	34	81.0	1006	4	US-09-023-905A-12	Sequence 12, Appl
35	34	81.0	1182	4	US-09-293-505-7	Sequence 7, Appli
36	34	81.0	1213	3	US-09-413-814-79	Sequence 79, Appl
37	33	78.6	10	3	US-08-602-999A-278	Sequence 278, App
38	33	78.6	10	4	US-09-500-124-278	Sequence 278, App
39	33	78.6	65	4	US-09-621-976-6387	Sequence 6387, Ap
40	33	78.6	95	4	US-09-314-268-132	Sequence 132, App
41	33	78.6	96	4	US-09-800-729-179	Sequence 179, App
42	33	78.6	112	4	US-09-621-976-6045	Sequence 6045, Ap
43	33	78.6	146	4	US-09-252-991A-24993	Sequence 24993, A
44	33	78.6	207	4	US-09-252-991A-29063	Sequence 29063, A
45	33	78.6	324	4	US-08-732-749-4	Sequence 4, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-641-803-1

; Sequence 1, Application US/09641803

; Patent No. 6500798

; GENERAL INFORMATION:

; APPLICANT: STANTON, G. John

; APPLICANT: HUGHES, Thomas K.

; APPLICANT: BOLDOGH, Istvan

; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND

; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS

; FILE REFERENCE: 265.00220101

; CURRENT APPLICATION NUMBER: US/09/641,803

; CURRENT FILING DATE: 2000-08-17

; PRIOR APPLICATION NUMBER: 60/149,310

; PRIOR FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: peptide  
US-09-641-803-1

Query Match 100.0%; Score 42; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | | | |  
Db 1 MQPPPLP 7

RESULT 2

US-08-339-214-16

; Sequence 16, Application US/08339214  
; Patent No. 6348334  
; GENERAL INFORMATION:  
; APPLICANT: Nagata, Shigikazu  
; APPLICANT: Suda, Takashi  
; APPLICANT: Takahashi, Tomoniro  
; APPLICANT: Nakamura, No. 6348334io  
; TITLE OF INVENTION: A Fas Ligand, A Fragment Thereof and DNA  
; TITLE OF INVENTION: Encoding the Same  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/339,214  
; FILING DATE: 10-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1110-139P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 278 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-339-214-16

Query Match 88.1%; Score 37; DB 4; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 60 QPPPLP 65

RESULT 3

US-08-339-214-26

; Sequence 26, Application US/08339214

; Patent No. 6348334

; GENERAL INFORMATION:

; APPLICANT: Nagata, Shigikazu

; APPLICANT: Suda, Takashi

; APPLICANT: Takahashi, Tomoniro

; APPLICANT: Nakamura, No. 6348334io

; TITLE OF INVENTION: A Fas Ligand, A Fragment Thereof and DNA

; TITLE OF INVENTION: Encoding the Same

; NUMBER OF SEQUENCES: 103

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch

; STREET: P.O. Box 747

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/339,214

; FILING DATE: 10-NOV-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy Jr., Gerald M.

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 1110-139P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-205-8000

; TELEFAX: 703-205-8050

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 278 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-339-214-26

Query Match 88.1%; Score 37; DB 4; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 60 QPPPLP 65

RESULT 4

US-09-131-237C-6  
 ; Sequence 6, Application US/09131237C  
 ; Patent No. 6599719  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Ni, Jian  
 ; APPLICANT: Rosen, Craig A.  
 ; TITLE OF INVENTION: Tumor Necrosis Factor Gamma  
 ; FILE REFERENCE: PF141P3  
 ; CURRENT APPLICATION NUMBER: US/09/131,237C  
 ; CURRENT FILING DATE: 1998-08-07  
 ; PRIOR APPLICATION NUMBER: 60/074,047  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 09/005,020  
 ; PRIOR FILING DATE: 1998-01-09  
 ; PRIOR APPLICATION NUMBER: 08/461,246  
 ; PRIOR FILING DATE: 1995-06-05  
 ; PRIOR APPLICATION NUMBER: PCT/US94/12880  
 ; PRIOR FILING DATE: 1994-11-07  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 6  
 ; LENGTH: 278  
 ; TYPE: PRT  
 ; ORGANISM: Rattus norvegicus  
 US-09-131-237C-6

Query Match 88.1%; Score 37; DB 4; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 60 QPPPLP 65

RESULT 5

US-09-252-991A-31536  
 ; Sequence 31536, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788



; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31536  
; LENGTH: 479  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31536

Query Match 88.1%; Score 37; DB 4; Length 479;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 193 QPPPLP 198

RESULT 6

US-09-252-991A-32104  
; Sequence 32104, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32104  
; LENGTH: 642  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32104

Query Match 88.1%; Score 37; DB 4; Length 642;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 41 QPPPLP 46

RESULT 7

US-08-718-661-2  
; Sequence 2, Application US/08718661  
; Patent No. 5876972  
; GENERAL INFORMATION:  
; APPLICANT:

```

; TITLE OF INVENTION: Nucleic acid molecules coding for mammalian
; TITLE OF INVENTION: tumor suppressor proteins and methods for their
isolation
; NUMBER OF SEQUENCES: 15
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,661
; FILING DATE:
; CLASSIFICATION: 530
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 667 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-661-2

```

```

Query Match          88.1%; Score 37; DB 2; Length 667;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches      6; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy      2 QPPPLP 7
        |||||
Db      527 QPPPLP 532

```

# RESULT 8

US-09-533-029-78

```

; Sequence 78, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78

```

; LENGTH: 721  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: G385  
US-09-533-029-78

Query Match 88.1%; Score 37; DB 4; Length 721;  
Best Local Similarity 85.7%; Pred. No. 4.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | |  
Db 205 MSPPPLP 211

RESULT 9

US-08-118-270-71

; Sequence 71, Application US/08118270  
; Patent No. 5508384  
; GENERAL INFORMATION:  
; APPLICANT: Murphy, Randall B.  
; APPLICANT: Schuster, David I.  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/118,270  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY=2A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 349 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-118-270-71

Query Match 85.7%; Score 36; DB 1; Length 349;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | |  
Db 290 MQPPPKP 296

RESULT 10

PCT-US93-08528-71

; Sequence 71, Application PC/TUS9308528  
; GENERAL INFORMATION:  
; APPLICANT: New York University  
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN  
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF  
; NUMBER OF SEQUENCES: 348  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/08528  
; FILING DATE: 09-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/943,236  
; FILING DATE: 10-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Townsend, Kevin G.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: MURPHY=2 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 349 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
PCT-US93-08528-71

Query Match 85.7%; Score 36; DB 5; Length 349;

Qy 1 MQPPPLP 7  
| | | | |  
Db 16 MQPPPPP 22

RESULT 12

PCT-US93-01112-34

; Sequence 34, Application PC/TUS9301112

; GENERAL INFORMATION:

; APPLICANT: Jamieson, Gordon A  
; APPLICANT: Dedman, John R  
; APPLICANT: Kaetzel, Marcia A  
; TITLE OF INVENTION: Calmodulin-Binding Peptides  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/01112  
; FILING DATE: 19930208  
; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/831,219  
; FILING DATE: 06-FEB-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Grant D  
; REGISTRATION NUMBER: 31,259  
; REFERENCE/DOCKET NUMBER: 272.001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (510) 601-2706  
; TELEFAX: (510) 655-3542

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 23 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US93-01112-34

Query Match 83.3%; Score 35; DB 5; Length 23;

Best Local Similarity 85.7%; Pred. No. 32;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | |  
Db 16 MQPPPPP 22

Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
      |||||  
Db 290 MQPPPKP 296

RESULT 11

US-08-268-251-34

; Sequence 34, Application US/08268251

; Patent No. 5585475

; GENERAL INFORMATION:

; APPLICANT: Jamieson, Gordon A

; APPLICANT: Dedman, John R

; APPLICANT: Kaetzel, Marcia A

; TITLE OF INVENTION: Calmodulin-Binding Peptides

; NUMBER OF SEQUENCES: 59

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Chiron Corporation

; STREET: 4560 Horton Street

; CITY: Emeryville

; STATE: CA

; COUNTRY: USA

; ZIP: 94608

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WordPerfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/268,251

; FILING DATE:

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/831,219

; FILING DATE: 06-FEB-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Green, Grant D

; REGISTRATION NUMBER: 31,259

; REFERENCE/DOCKET NUMBER: 272.001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (510) 601-2706

; TELEFAX: (510) 655-3542

; INFORMATION FOR SEQ ID NO: 34:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 23 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-268-251-34

Query Match 83.3%; Score 35; DB 1; Length 23;

Best Local Similarity 85.7%; Pred. No. 32;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 13

US-09-388-349-9

```
; Sequence 9, Application US/09388349
; Patent No. 6365370
; GENERAL INFORMATION:
; APPLICANT: Dubin, Adreinne E
; APPLICANT: Erlander, Mark G
; APPLICANT: Huvar, Arne
; APPLICANT: Huvar, Rene
; APPLICANT: Buehler, Lukas K
; TITLE OF INVENTION: DNA Encoding A Human Subunit 5-HT3-C of the 5-HT3
Serotonin Receptor
; FILE REFERENCE: ORT-1039
; CURRENT APPLICATION NUMBER: US/09/388,349
; CURRENT FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-388-349-9
```

```
Query Match          83.3%; Score 35; DB 4; Length 447;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches      5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 QPPPLP 7
        ||||:|
Db      339 QPPMP 344
```

RESULT 14

US-09-096-731A-14

```
; Sequence 14, Application US/09096731A
; Patent No. 6489454
; GENERAL INFORMATION:
; APPLICANT: Liu, Qiong
; APPLICANT: Hengartner, Michael O.
; APPLICANT: Bogaert, Thierry Andre Oliver Eddy
; APPLICANT: Van Criekeing, Wim Maria Rene
; TITLE OF INVENTION: ENGULFMENT GENE AND USES THEREOF
; FILE REFERENCE: CSHL97-07pA
; CURRENT APPLICATION NUMBER: US/09/096,731A
; CURRENT FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: US 60/072,324
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; OTHER INFORMATION: Proline/Serine Rich Region
US-09-096-731A-14
```

Query Match 81.0%; Score 34; DB 4; Length 76;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
| | | | : |  
Db 38 QPPVP 43

RESULT 15

US-09-252-991A-24956  
; Sequence 24956, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24956  
; LENGTH: 158  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24956

Query Match 81.0%; Score 34; DB 4; Length 158;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
: | | | | |  
Db 136 LSPPPLP 142

Search completed: August 24, 2004, 15:55:10  
Job time : 9.6791 secs



OM protein - protein search, using sw model

Run on: August 24, 2004, 15:26:28 ; Search time 6.79105 Seconds  
(without alignments)  
99.151 Million cell updates/sec

Title: US-09-641-801-1  
Perfect score: 42  
Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	40	95.2	1776	2	G86280	protein T5E21.13 [
2	39	92.9	612	2	E82756	beta-galactosidase
3	38	90.5	536	2	T02441	DNA-(apurinic or a
4	38	90.5	542	2	H86239	protein F20B24.8 [
5	37	88.1	164	2	T15525	hypothetical prote
6	37	88.1	178	2	T00644	hypothetical prote
7	37	88.1	186	2	T22685	hypothetical prote
8	37	88.1	211	2	S55129	transcription fact
9	37	88.1	278	2	A49266	fas ligand - rat
10	37	88.1	393	2	T00575	probable WRKY-type
11	37	88.1	471	2	T21349	hypothetical prote
12	37	88.1	483	2	A25896	beta-adrenergic re
13	37	88.1	588	2	E96633	probable Serine/Th

14	37	88.1	620	2	S06733	hydroxyproline-ric
15	37	88.1	649	2	JN0809	drebrin E (clone g
16	37	88.1	651	2	T14763	hypothetical prote
17	37	88.1	731	2	JC7701	ARHGAP9 protein -
18	37	88.1	749	2	G86186	hypothetical prote
19	37	88.1	830	2	T18860	hypothetical prote
20	37	88.1	1124	2	JX0293	zinc finger protei
21	37	88.1	1125	2	T19193	hypothetical prote
22	37	88.1	1154	2	A56242	E-box-binding repr
23	37	88.1	1291	2	T17242	hypothetical prote
24	37	88.1	1364	2	T00250	MEGF2 protein - hu
25	37	88.1	1560	2	T00080	hypothetical prote
26	36	85.7	489	2	S23410	FUN19 protein - ye
27	36	85.7	505	2	S72273	actin-depolymerizi
28	36	85.7	542	2	JC7391	Ca2+-binding prote
29	36	85.7	610	2	T22687	hypothetical prote
30	36	85.7	736	2	I51691	dishevelled homolo
31	36	85.7	795	2	T49835	hypothetical prote
32	36	85.7	823	2	A36378	probable transcrip
33	35	83.3	267	2	T47307	hypothetical prote
34	35	83.3	300	2	T25119	hypothetical prote
35	35	83.3	312	2	A86279	Fl4L17.21 protein
36	35	83.3	321	2	G86411	protein FlK23.14 [
37	35	83.3	337	2	F96507	hypothetical prote
38	35	83.3	414	2	A88485	protein F23F12.6 [
39	35	83.3	426	2	T47365	hypothetical prote
40	35	83.3	486	2	T21481	hypothetical prote
41	35	83.3	494	2	T19550	hypothetical prote
42	35	83.3	548	2	S52735	CW17R protein - mo
43	35	83.3	587	2	S49942	hypothetical prote
44	35	83.3	597	2	S72468	probable transcrip
45	35	83.3	639	2	G02919	transcription fact

## ALIGNMENTS

### RESULT 1

G86280

protein T5E21.13 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: G86280

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86280  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1776 <STO>  
A;Cross-references: GB:AE005172; NID:g7527720; PIDN:AAF63169.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: T5E21.13  
A;Map position: 1

Query Match 95.2%; Score 40; DB 2; Length 1776;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
| | | | | : |  
Db 1643 MQPPMP 1649

## RESULT 2

E82756

beta-galactosidase XF0840 [imported] - *Xylella fastidiosa* (strain 9a5c)

C;Species: *Xylella fastidiosa*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C;Accession: E82756

R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil.

Nature 406, 151-157, 2000

A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A;Reference number: A82515; MUID:20365717; PMID:10910347

A;Note: for a complete list of authors see reference number A59328 below

A;Accession: E82756

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-612 <SIM>

A;Cross-references: GB:AE003923; GB:AE003849; NID:g9105736; PIDN:AAF83650.1; GSPDB:GN00128; XFSC:XF0840

A;Experimental source: strain 9a5c

R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alves, L.M.C.; Araya, J.E.; Baia, G.S.; Baptista, C.S.; Barros, M.H.; Bonaccorsi, E.D.; Bordin, S.; Bove, J.M.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.; Colauto, N.B.; Colombo, C.; Costa, F.F.; Costa, M.C.R.; Costa-Neto, C.M.; Coutinho, L.L.; Cristofani, M.; Dias-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme, M.; Furlan, L.R.; Garnier, M.; Goldman, G.H.; Goldman, M.H.S.; Gomes, S.L.; Gruber, A.; Ho, P.L.; Hoheisel, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, F.; Lambais, M.R.; Leite, L.C.C.; Lemos, E.G.M.; Lemos, M.V.F.; Lopes, S.A.; Lopes, C.R.; Machado,

J.A.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.L.  
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Monteiro-Vitorello, C.B.; Moon, D.H.; Nagai, M.A.; Nascimento, A.L.T.O.; Netto, L.E.S.; Nhani Jr., A.; Nobrega, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Paris, A.; Peixoto, B.R.; Pereira, G.A.G.; Pereira Jr., H.A.; Pesquero, J.B.; Quaggio, R.B.; Roberto, P.G.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki, H.E.  
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, J.F.; Silvestri, M.L.Z.; Siqueira, W.J.; de Souza, A.A.; de Souza, A.P.; Terenzi, M.F.; Truffi, D.; Tsai, S.M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zago, M.A.; Zatz, M.; Meidanis, J.; Setubal, J.C.  
 A;Reference number: A59328  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: XF0840  
 C;Superfamily: beta-galactosidase bga

Query Match 92.9%; Score 39; DB 2; Length 612;  
 Best Local Similarity 85.7%; Pred. No. 98;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQPPPLP 7  
 :|||||  
 Db 356 LQPPPLP 362

# RESULT 3

T02441

DNA-(apurinic or apyrimidinic site) lyase (ARP) [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T26J13.5

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 02-Feb-2001

C;Accession: T02441; A84842

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.

submitted to the EMBL Data Library, June 1998

A;Description: Arabidopsis thaliana chromosome II BAC T26J13 genomic sequence.

A;Reference number: Z14673

A;Accession: T02441

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-536 <ROU>

A;Cross-references: EMBL:AC004625; NID:g3241939; PID:g3241944

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: A84842  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-536 <STO>  
A;Cross-references: GB:AE002093; NID:g3241944; PIDN:AAC23731.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g41460; T26J13.5  
A;Map position: 2  
A;Introns: 18/2; 78/3; 110/2; 137/1; 210/1; 317/3; 354/3; 400/3; 416/3; 443/3; 497/1

Query Match 90.5%; Score 38; DB 2; Length 536;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|:|||||  
Db 264 MRPPPLP 270

#### RESULT 4

H86239

protein F20B24.8 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: H86239

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pal, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: H86239

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-542 <STO>

A;Cross-references: GB:AE005172; NID:g6573750; PIDN:AAF17670.1; GSPDB:GN00141

C;Genetics:

A;Gene: F20B24.8

A;Map position: 1

Query Match 90.5%; Score 38; DB 2; Length 542;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 105 VQPPPLP 111

RESULT 5

T15525

hypothetical protein C16B8.3 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000

C;Accession: T15525

R;Bentley, D.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of *C. elegans* cosmid C16B8.

A;Reference number: Z18365

A;Accession: T15525

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-164 <BEN>

A;Cross-references: EMBL:U41031; NID:g1098982; PID:g1098985; PIDN:AAA82620.1;

CESP:C16B8.3

C;Genetics:

A;Gene: CESP:C16B8.3

C;Superfamily: proline-rich protein

Query Match 88.1%; Score 37; DB 2; Length 164;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 97 QPPPLP 102

RESULT 6

T00644

hypothetical protein F3I6.7 - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999

C;Accession: T00644

R;Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.;

Araujo, R.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Oji, O.; Osborne,

B.I.; Shinn, P.; Sun, H.; Toriumi, M.; Vysotskaia, V.S.; Yu, G.; Ecker, J.;

Theologis, A.; Davis, R.W.

submitted to the EMBL Data Library, February 1998

A;Reference number: Z14197

A;Accession: T00644

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-178 <FED>

A;Cross-references: EMBL:AC002396; NID:g2749918; PID:g2829866; GSPDB:GN00059;

ATSP:F3I6.7

C;Genetics:

A;Gene: ATSP:F3I6.7  
A;Map position: 1  
A;Introns: 50/3; 87/1  
C;Superfamily: Arabidopsis thaliana hypothetical protein F3I6.7

Query Match 88.1%; Score 37; DB 2; Length 178;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||  
Db 162 QPPPLP 167

RESULT 7

T22685

hypothetical protein F55A11.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T22685

R;Kershaw, J.

submitted to the EMBL Data Library, May 1996

A;Reference number: Z19600

A;Accession: T22685

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-186 <WIL>

A;Cross-references: EMBL:Z72511; PIDN:CAA96655.1; GSPDB:GN00023; CESP:F55A11.1

A;Experimental source: clone F55A11

C;Genetics:

A;Gene: CESP:F55A11.1

A;Map position: 5

A;Introns: 33/3; 59/2; 146/2

Query Match 88.1%; Score 37; DB 2; Length 186;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||  
Db 127 QPPPLP 132

RESULT 8

S55129

transcription factor RGM1 - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YM8010.12c; protein YMR182c

C;Species: *Saccharomyces cerevisiae*

C;Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 21-Jul-2000

C;Accession: S55129; S17249

R;Churcher, C.M.

submitted to the EMBL Data Library, June 1995

A;Reference number: S55118

A;Accession: S55129

A;Molecule type: DNA

A;Residues: 1-211 <CHU>

A;Cross-references: EMBL:Z49808; NID:g854440; PID:g854452; MIPS:YMR182c

A;Experimental source: strain AB972  
 R;Estruch, F.  
 Nucleic Acids Res. 19, 4873-4877, 1991  
 A;Title: The yeast putative transcriptional repressor RGM1 is a proline-rich zinc finger protein.  
 A;Reference number: S17249; MUID:92020118; PMID:1923755  
 A;Accession: S17249  
 A;Molecule type: DNA  
 A;Residues: 1-113,'V',115-211 <EST>  
 A;Cross-references: EMBL:X59861; NID:g4317; PIDN:CAA42521.1; PID:g4318  
 C;Genetics:  
 A;Gene: SGD:RGM1  
 A;Cross-references: SGD:S0004794; MIPS:YMR182c  
 A;Map position: 13R  
 C;Keywords: DNA binding; transcription factor; zinc finger

Query Match 88.1%; Score 37; DB 2; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 154 QPPPLP 159

#### RESULT 9

A49266

fas ligand - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999

C;Accession: A49266

R;Suda, T.; Takahashi, T.; Golstein, P.; Nagata, S.

Cell 75, 1169-1178, 1993

A;Title: Molecular cloning and expression of the Fas ligand, a novel member of the tumor necrosis factor family.

A;Reference number: A49266; MUID:94084792; PMID:7505205

A;Accession: A49266

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-278 <SUD>

A;Cross-references: GB:U03470; NID:g440178; PIDN:AAC52129.1; PID:g440179

C;Keywords: glycoprotein; transmembrane protein

Query Match 88.1%; Score 37; DB 2; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 60 QPPPLP 65

#### RESULT 10

T00575

probable WRKY-type DNA binding protein [imported] - Arabidopsis thaliana

N;Alternate names: probable DNA-binding protein T27E13.1

C;Species: Arabidopsis thaliana (mouse-ear cress)



C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 23-Mar-2001  
 C;Accession: T00575; B84706  
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.  
 submitted to the EMBL Data Library, May 1998  
 A;Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.  
 A;Reference number: Z14178  
 A;Accession: T00575  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-393 <ROU>  
 A;Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150397  
 A;Experimental source: cultivar Columbia  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: B84706  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-393 <STO>  
 A;Cross-references: GB:AE002093; NID:g2347191; PIDN:AAC16930.1; GSPDB:GN00139  
 C;Genetics:  
 A;Gene: At2g30250; T27E13.1  
 A;Map position: 2  
 A;Introns: 79/3; 117/3; 294/2; 348/2

Query Match 88.1%; Score 37; DB 2; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QPPPLP 7  
 | | | | |  
 Db 50 QPPPLP 55

RESULT 11  
 T21349  
 hypothetical protein F25H2.8 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C;Accession: T21349  
 R;Wilkinson, J.  
 submitted to the EMBL Data Library, September 1996  
 A;Reference number: Z19411  
 A;Accession: T21349  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-471 <WIL>

A;Cross-references: EMBL:Z79754; PIDN:CAB02096.1; GSPDB:GN00019; CESP:F25H2.8  
A;Experimental source: clone F25H2  
C;Genetics:  
A;Gene: CESP:F25H2.8  
A;Map position: 1  
A;Introns: 26/3; 138/3; 302/3; 353/3; 379/2  
C;Superfamily: Caenorhabditis elegans hypothetical protein F25H2.8

Query Match 88.1%; Score 37; DB 2; Length 471;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:||||:|  
Db 451 LQPPPIP 457

RESULT 12

A25896

beta-adrenergic receptor - turkey

C;Species: Meleagris gallopavo (common turkey)

C;Date: 16-Aug-1988 #sequence\_revision 16-Aug-1988 #text\_change 13-Aug-1999

C;Accession: A25896

R;Yarden, Y.; Rodriguez, H.; Wong, S.K.F.; Brandt, D.R.; May, D.C.; Burnier, J.;  
Harkins, R.N.; Chen, E.Y.; Ramachandran, J.; Ullrich, A.; Ross, E.M.  
Proc. Natl. Acad. Sci. U.S.A. 83, 6795-6799, 1986

A;Title: The avian beta-adrenergic receptor: primary structure and membrane topology.

A;Reference number: A25896; MUID:86313664; PMID:3018746

A;Accession: A25896

A;Molecule type: mRNA

A;Residues: 1-483 <YAR>

A;Cross-references: GB:M14379; NID:g213891; PIDN:AAA49627.1; PID:g213892

A;Experimental source: erythrocyte

C;Superfamily: vertebrate rhodopsin

C;Keywords: G protein-coupled receptor; transmembrane protein

Query Match 88.1%; Score 37; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||||  
Db 256 QPPPLP 261

RESULT 13

E96633

probable Serine/Threonine protein kinase F8A5.31 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: E96633

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;

Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
 Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
 Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
 Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
 S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzialli, A.; Militscher, J.; Miranda,  
 M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
 Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
 S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
 Venter, J.C.; Davis, R.W.  
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A;Reference number: A86141; MUID:21016719; PMID:11130712  
 A;Accession: E96633  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-588 <STO>  
 A;Cross-references: GB:AE005173; NID:g2462749; PIDN:AAB71968.1; GSPDB:GN00141  
 C;Genetics:  
 A;Gene: F8A5.31  
 A;Map position: 1

Query Match 88.1%; Score 37; DB 2; Length 588;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 550 QPPPLP 555

RESULT 14  
 S06733  
 hydroxyproline-rich glycoprotein precursor - common tobacco  
 C;Species: Nicotiana tabacum (common tobacco)  
 C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jul-2000  
 C;Accession: S06733  
 R;Keller, B.; Lamb, C.J.  
 Genes Dev. 3, 1639-1646, 1989  
 A;Title: Specific expression of a novel cell wall hydroxyproline-rich  
 glycoprotein gene in lateral root initiation.  
 A;Reference number: S06733; MUID:90128263; PMID:2612909  
 A;Accession: S06733  
 A;Molecule type: DNA  
 A;Residues: 1-620 <KEL>  
 A;Cross-references: EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867  
 C;Superfamily: hydroxyproline-rich glycoprotein  
 C;Keywords: glycoprotein

Query Match 88.1%; Score 37; DB 2; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||

Db 425 QPPPLP 430

RESULT 15

JN0809

drebrin E (clone gDbh13) - human

C;Species: Homo sapiens (man)

C;Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 21-Jul-2000

C;Accession: JN0809

R;Toda, M.; Shirao, T.; Minoshima, S.; Shimizu, N.; Toya, S.; Uyemura, K.

Biochem. Biophys. Res. Commun. 196, 468-472, 1993

A;Title: Molecular cloning of cDNA encoding human drebrin E and chromosomal mapping of its gene.

A;Reference number: JN0809; MUID:94030036; PMID:8216329

A;Accession: JN0809

A;Molecule type: mRNA

A;Residues: 1-649 <TOD>

A;Cross-references: DDBJ:D17530; NID:g498650; PIDN:BAA04480.1; PID:g498651

A;Experimental source: brain

C;Comment: This protein is an actin-binding protein.

C;Genetics:

A;Map position: 5

C;Keywords: actin binding

Query Match 88.1%; Score 37; DB 2; Length 649;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7

|||||

Db 363 QPPPLP 368

Search completed: August 24, 2004, 15:52:38

Job time : 10.791 secs

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:51:19 ; Search time 25.3358 Seconds  
 (without alignments)  
 86.825 Million cell updates/sec

Title: US-09-641-801-1  
 Perfect score: 42  
 Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Score	Match	Length	DB	ID	Description
No.						

1	42	100.0	7	14	US-10-281-652-1	Sequence 1, Appli
2	39	92.9	76	12	US-10-424-599-147458	Sequence 147458,
3	39	92.9	132	9	US-09-864-761-37187	Sequence 37187, A
4	39	92.9	332	16	US-10-437-963-143948	Sequence 143948,
5	39	92.9	407	16	US-10-437-963-175374	Sequence 175374,
6	39	92.9	572	15	US-10-369-493-9267	Sequence 9267, Ap
7	39	92.9	572	15	US-10-369-493-9495	Sequence 9495, Ap
8	39	92.9	612	15	US-10-369-493-17539	Sequence 17539, A
9	38	90.5	189	12	US-10-424-599-148127	Sequence 148127,
10	38	90.5	256	16	US-10-408-765A-2896	Sequence 2896, Ap
11	38	90.5	414	12	US-10-424-599-147426	Sequence 147426,
12	38	90.5	812	16	US-10-437-963-143219	Sequence 143219,
13	37	88.1	38	10	US-09-899-495-68	Sequence 68, Appl
14	37	88.1	38	10	US-09-802-668-109	Sequence 109, App
15	37	88.1	38	14	US-10-243-475-109	Sequence 109, App
16	37	88.1	61	12	US-10-424-599-193704	Sequence 193704,
17	37	88.1	66	12	US-10-424-599-170169	Sequence 170169,
18	37	88.1	72	16	US-10-437-963-177968	Sequence 177968,
19	37	88.1	75	16	US-10-437-963-159895	Sequence 159895,
20	37	88.1	82	16	US-10-437-963-145289	Sequence 145289,
21	37	88.1	83	16	US-10-437-963-190127	Sequence 190127,
22	37	88.1	91	16	US-10-437-963-160130	Sequence 160130,
23	37	88.1	98	12	US-10-425-114-48917	Sequence 48917, A
24	37	88.1	98	16	US-10-437-963-137655	Sequence 137655,
25	37	88.1	104	16	US-10-437-963-122299	Sequence 122299,
26	37	88.1	112	16	US-10-437-963-150616	Sequence 150616,
27	37	88.1	113	16	US-10-437-963-122527	Sequence 122527,
28	37	88.1	114	12	US-10-424-599-236948	Sequence 236948,
29	37	88.1	115	16	US-10-437-963-194122	Sequence 194122,
30	37	88.1	118	16	US-10-437-963-196882	Sequence 196882,
31	37	88.1	129	12	US-10-424-599-192878	Sequence 192878,
32	37	88.1	144	12	US-10-424-599-264209	Sequence 264209,
33	37	88.1	146	12	US-10-424-599-270730	Sequence 270730,
34	37	88.1	147	16	US-10-437-963-111116	Sequence 111116,
35	37	88.1	159	16	US-10-437-963-131050	Sequence 131050,
36	37	88.1	162	16	US-10-437-963-203781	Sequence 203781,
37	37	88.1	163	16	US-10-437-963-155754	Sequence 155754,
38	37	88.1	164	16	US-10-437-963-179013	Sequence 179013,
39	37	88.1	167	16	US-10-437-963-130977	Sequence 130977,
40	37	88.1	171	16	US-10-437-963-203782	Sequence 203782,
41	37	88.1	184	16	US-10-437-963-116550	Sequence 116550,
42	37	88.1	187	16	US-10-437-963-185697	Sequence 185697,
43	37	88.1	189	12	US-10-424-599-225072	Sequence 225072,
44	37	88.1	191	16	US-10-437-963-175982	Sequence 175982,
45	37	88.1	194	16	US-10-437-963-146891	Sequence 146891,

#### ALIGNMENTS

RESULT 1  
 US-10-281-652-1  
 ; Sequence 1, Application US/10281652  
 ; Publication No. US20030091606A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: STANTON, G. John

```
; APPLICANT: HUGHES, Thomas K.
; APPLICANT: BOLDOGH, Istvan
; TITLE OF INVENTION: USE OF COLOSTRININ, CONSTITUENT PEPTIDES THEREOF, AND
; TITLE OF INVENTION: ANALOGS THEREOF AS OXIDATIVE STRESS REGULATORS
; FILE REFERENCE: 265.00220101
; CURRENT APPLICATION NUMBER: US/10/281,652
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/641,803
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,310
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: peptide
US-10-281-652-1
```

```
Query Match          100.0%; Score 42; DB 14; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches      7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MQPPPLP 7
        |||||
Db      1 MQPPPLP 7
```

## RESULT 2

```
US-10-424-599-147458
; Sequence 147458, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 147458
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_104173C.1.pep
US-10-424-599-147458
```

```
Query Match          92.9%; Score 39; DB 12; Length 76;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
```

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MQPPPLP 7  
:|||||  
Db 19 LQPPPLP 25

RESULT 3

US-09-864-761-37187

; Sequence 37187, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES  
USEFUL FOR

; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203



; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 37187  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL096814.11  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.3  
; OTHER INFORMATION: EST\_HUMAN HIT: AW833678.1, EVALUE 4.00e-03  
US-09-864-761-37187

Query Match 92.9%; Score 39; DB 9; Length 132;  
Best Local Similarity 85.7%; Pred. No. 3.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 116 LQPPPLP 122

RESULT 4

US-10-437-963-143948  
; Sequence 143948, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 143948  
; LENGTH: 332  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_44807C.1.pep

US-10-437-963-143948

Query Match 92.9%; Score 39; DB 16; Length 332;  
Best Local Similarity 85.7%; Pred. No. 8.7e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 85 LQPPPLP 91

RESULT 5

US-10-437-963-175374

; Sequence 175374, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 175374  
; LENGTH: 407  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(407)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_73225C.1.pep  
US-10-437-963-175374

Query Match 92.9%; Score 39; DB 16; Length 407;  
Best Local Similarity 85.7%; Pred. No. 1e+03;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 105 LQPPPLP 111

RESULT 6

US-10-369-493-9267

; Sequence 9267, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9267
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9267
```

```
Query Match          92.9%;  Score 39;  DB 15;  Length 572;
Best Local Similarity 85.7%;  Pred. No. 1.4e+03;
Matches      6;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1 MQPPPLP 7
        :|||||
Db      356 LQPPPLP 362
```

#### RESULT 7

```
US-10-369-493-9495
; Sequence 9495, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9495
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9495
```

```
Query Match          92.9%;  Score 39;  DB 15;  Length 572;
Best Local Similarity 85.7%;  Pred. No. 1.4e+03;
Matches      6;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;
```

Qy 1 MQPPPLP 7  
:|||||  
Db 356 LQPPPLP 362

RESULT 8

US-10-369-493-17539

; Sequence 17539, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 17539

; LENGTH: 612

; TYPE: PRT

; ORGANISM: Xylella fastidiosa

US-10-369-493-17539

Query Match 92.9%; Score 39; DB 15; Length 612;

Best Local Similarity 85.7%; Pred. No. 1.5e+03;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 356 LQPPPLP 362

RESULT 9

US-10-424-599-148127

; Sequence 148127, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 148127

; LENGTH: 189  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(189)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_104780C.1.pep  
US-10-424-599-148127

Query Match 90.5%; Score 38; DB 12; Length 189;  
Best Local Similarity 85.7%; Pred. No. 7.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|:|||||  
Db 105 MKPPPLP 111

RESULT 10

US-10-408-765A-2896  
; Sequence 2896, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2896  
; LENGTH: 256  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-2896

Query Match 90.5%; Score 38; DB 16; Length 256;  
Best Local Similarity 85.7%; Pred. No. 9.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 191 VQPPPLP 197

RESULT 11

US-10-424-599-147426  
; Sequence 147426, Application US/10424599

; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 147426  
; LENGTH: 414  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_104144C.1.pep  
US-10-424-599-147426

Query Match 90.5%; Score 38; DB 12; Length 414;  
Best Local Similarity 85.7%; Pred. No. 1.4e+03;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|:|||||  
Db 105 MKPPPLP 111

RESULT 12

US-10-437-963-143219  
; Sequence 143219, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 143219  
; LENGTH: 812  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_44149C.1.pep  
US-10-437-963-143219

Query Match 90.5%; Score 38; DB 16; Length 812;  
Best Local Similarity 85.7%; Pred. No. 2.5e+03;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 26 VQPPPLP 32

RESULT 13

US-09-899-495-68  
; Sequence 68, Application US/09899495  
; Publication No. US20030088060A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Christopher W.  
; APPLICANT: Roberts, Steven L.  
; APPLICANT: Karnovsky, Alla M.  
; APPLICANT: Ruble, Cara L.  
; TITLE OF INVENTION: Human Ion Channels  
; FILE REFERENCE: 00188US1  
; CURRENT APPLICATION NUMBER: US/09/899,495  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 60/215,815  
; PRIOR FILING DATE: 2000-07-05  
; PRIOR APPLICATION NUMBER: 60/216,481  
; PRIOR FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: 60/216,479  
; PRIOR FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: 60/216,482  
; PRIOR FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: 60/217,096  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 68  
; LENGTH: 38  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-899-495-68

Query Match 88.1%; Score 37; DB 10; Length 38;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 25 QPPPLP 30

RESULT 14

US-09-802-668-109  
; Sequence 109, Application US/09802668  
; Publication No. US20030190714A1  
; GENERAL INFORMATION:  
; APPLICANT: Wood, Linda  
; APPLICANT: Vogeli, Gabriel

```
; APPLICANT: Karnovsky, Alla
; APPLICANT: Linske-O'Connell, Lisa I.
; APPLICANT: Wang, Jun
; APPLICANT: Liu, Derong
; TITLE OF INVENTION: Human Ion Channels
; FILE REFERENCE: 00069US1
; CURRENT APPLICATION NUMBER: US/09/802,668
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,517
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
;   LENGTH: 38
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-802-668-109
```

```
Query Match          88.1%;   Score 37;   DB 10;   Length 38;
Best Local Similarity 100.0%;   Pred. No. 2.4e+02;
Matches      6;   Conservative      0;   Mismatches      0;   Indels      0;   Gaps      0;
```

```
Qy          2 QPPPLP 7
           |||||
Db          25 QPPPLP 30
```

# RESULT 15

```
US-10-243-475-109
; Sequence 109, Application US/10243475
; Publication No. US20030194720A1
; GENERAL INFORMATION:
; APPLICANT: Roberds, Steven L.
; APPLICANT: Benjamin, Christopher W.
; APPLICANT: Karnovsky, Alla M.
; APPLICANT: Ruble, Cara L.
; TITLE OF INVENTION: Human Ion Channels
; FILE REFERENCE: PHRM0039-100
; CURRENT APPLICATION NUMBER: US/10/243,475
; CURRENT FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: 60/403,254
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 60/318,733
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 109
;   LENGTH: 38
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-243-475-109
```

```
Query Match          88.1%;   Score 37;   DB 14;   Length 38;
Best Local Similarity 100.0%;   Pred. No. 2.4e+02;
Matches      6;   Conservative      0;   Mismatches      0;   Indels      0;   Gaps      0;
```

```
Qy          2 QPPPLP 7
```



Db            |||||  
             25 QPPPLP 30

Search completed: August 24, 2004, 16:41:11  
Job time : 28.3358 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 24, 2004, 15:23:00 ; Search time 21.6269 Seconds  
(without alignments)  
102.124 Million cell updates/sec

Title: US-09-641-801-1  
Perfect score: 42  
Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

§  
Result Query  
No. Score Match Length DB ID Description  
-----

1	42	100.0	261	5	Q9VAX5	Q9vax5 drosophila
2	40	95.2	600	2	Q8GPV8	Q8gpv8 pseudomonas
3	40	95.2	785	10	Q8RXF1	Q8rxfl1 arabidopsis
4	40	95.2	1776	10	Q9MA20	Q9ma20 arabidopsis
5	39	92.9	612	16	Q9PF38	Q9pf38 xylella fas
6	39	92.9	612	16	Q87AJ2	Q87aj2 xylella fas
7	39	92.9	1404	5	Q9GRR6	Q9grr6 leishmania
8	38	90.5	542	10	Q9SGY5	Q9sgy5 arabidopsis
9	38	90.5	678	4	Q8TES7	Q8tes7 homo sapien
10	38	90.5	812	10	Q8LIE4	Q8lie4 oryza sativ
11	38	90.5	2165	5	Q8T1F5	Q8t1f5 dictyosteli
12	38	90.5	2879	5	Q8IC57	Q8ic57 plasmodium
13	37	88.1	44	8	Q85G37	Q85g37 cyanidiosch
14	37	88.1	72	15	Q87610	Q87610 chimpanzee
15	37	88.1	73	15	Q87603	Q87603 chimpanzee
16	37	88.1	99	15	Q87482	Q87482 chimpanzee
17	37	88.1	126	5	Q9NDS3	Q9nds3 penaeus jap
18	37	88.1	126	5	Q9NDS2	Q9nds2 penaeus jap
19	37	88.1	128	10	Q8GYP2	Q8gyp2 arabidopsis
20	37	88.1	137	10	Q84QQ0	Q84qq0 oryza sativ
21	37	88.1	164	5	Q18043	Q18043 caenorhabdi
22	37	88.1	178	10	Q48681	Q48681 arabidopsis
23	37	88.1	186	5	Q20796	Q20796 caenorhabdi
24	37	88.1	190	11	Q811B5	Q811b5 mus musculu
25	37	88.1	222	10	Q8S1H5	Q8s1h5 oryza sativ
26	37	88.1	223	16	Q8PI93	Q8pi93 xanthomonas
27	37	88.1	223	16	Q8P6X9	Q8p6x9 xanthomonas
28	37	88.1	226	10	Q7XDT9	Q7xdt9 oryza sativ
29	37	88.1	276	16	Q8DHT8	Q8dht8 synechococc
30	37	88.1	294	10	Q9LG53	Q9lg53 oryza sativ
31	37	88.1	302	4	Q86U16	Q86u16 homo sapien
32	37	88.1	322	4	Q9BSM2	Q9bsm2 homo sapien
33	37	88.1	325	16	Q89MW5	Q89mw5 bradyrhizob
34	37	88.1	346	11	Q99LN2	Q99ln2 mus musculu
35	37	88.1	353	11	Q8BTT8	Q8btt8 mus musculu
36	37	88.1	357	4	Q96RZ4	Q96rz4 homo sapien
37	37	88.1	358	4	Q8N249	Q8n249 homo sapien
38	37	88.1	401	4	Q8IVY2	Q8ivy2 homo sapien
39	37	88.1	420	16	Q7U607	Q7u607 synechococc
40	37	88.1	421	10	Q9FTS7	Q9fts7 oryza sativ
41	37	88.1	423	10	Q8H5B0	Q8h5b0 oryza sativ
42	37	88.1	431	5	Q8SXY7	Q8sxy7 drosophila
43	37	88.1	440	5	Q9V6M1	Q9v6m1 drosophila
44	37	88.1	456	5	Q9N9U6	Q9n9u6 leishmania
45	37	88.1	471	4	Q7Z6B2	Q7z6b2 homo sapien

#### ALIGNMENTS

##### RESULT 1

Q9VAX5

ID Q9VAX5 PRELIMINARY; PRT; 261 AA.

AC Q9VAX5;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE CG12872 protein.  
 GN CG12872.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanenvong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003763; AAF56772.2; -.  
 DR FlyBase; FBgn0039569; CG12872.  
 SQ SEQUENCE 261 AA; 28401 MW; D5753D823160DAA6 CRC64;

Query Match 100.0%; Score 42; DB 5; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||  
 Db 194 MQPPPLP 200

## RESULT 2

Q8GPV8

ID Q8GPV8 PRELIMINARY; PRT; 600 AA.  
 AC Q8GPV8;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SG17M;  
 RX MEDLINE=22313472; PubMed=12426355;

RA Larbig K.D., Christmann A., Johann A., Klockgether J., Hartsch T.,  
 RA Merkl R., Wiehlmann L., Fritz H.J., Tummeler B.;  
 RT "Gene Islands Integrated into tRNA(Gly) Genes Confer Genome Diversity  
 RT on a Pseudomonas aeruginosa Clone.";  
 RL J. Bacteriol. 184:6665-6680(2002).  
 DR EMBL; AF440524; AAN62266.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 600 AA; 66371 MW; BF43C0A0E10FFF9D CRC64;

Query Match 95.2%; Score 40; DB 2; Length 600;  
 Best Local Similarity 85.7%; Pred. No. 65;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 |||||:  
 Db 460 MQPPMP 466

# RESULT 3

## Q8RXF1

ID Q8RXF1 PRELIMINARY; PRT; 785 AA.  
 AC Q8RXF1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Splicing factor, putative.  
 GN AT1G14650.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY081293; AAL91182.1; -.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0006396; P:RNA processing; IEA.  
 DR InterPro; IPR000061; Surp.  
 DR InterPro; IPR000626; Ubiquitin.  
 DR Pfam; PF01805; Surp; 2.  
 DR Pfam; PF00240; ubiquitin; 1.  
 DR SMART; SM00648; SWAP; 2.  
 DR SMART; SM00213; UBQ; 1.  
 DR PROSITE; PS50128; SURP; 2.  
 DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
 SQ SEQUENCE 785 AA; 87536 MW; A1958C193E63BDC4 CRC64;

Query Match 95.2%; Score 40; DB 10; Length 785;  
 Best Local Similarity 85.7%; Pred. No. 84;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|||||:  
Db 652 MQPPPMP 658

RESULT 4

Q9MA20

ID Q9MA20 PRELIMINARY; PRT; 1776 AA.  
AC Q9MA20;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE T5E21.13.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Tambunga G., Altafi H., Bei Q., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Federspiel N.A.,  
RA Theologis A., Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC T5E21 from chromosome  
RT I.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AC010657; AAF63169.1; -.  
DR PIR; G86280; G86280.  
DR HSSP; P02248; 1UBI.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0006396; P:RNA processing; IEA.  
DR InterPro; IPR000061; Surp.  
DR InterPro; IPR000626; Ubiquitin.  
DR Pfam; PF01805; Surp; 4.  
DR Pfam; PF00240; ubiquitin; 1.  
DR SMART; SM00648; SWAP; 4.  
DR SMART; SM00213; UBQ; 1.  
DR PROSITE; PS50128; SURP; 4.  
DR PROSITE; PS50053; UBIQUITIN\_2; 1.  
SQ SEQUENCE 1776 AA; 200270 MW; C1C28169934E0A27 CRC64;

Query Match 95.2%; Score 40; DB 10; Length 1776;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
|||||:  
Db 1643 MQPPPMP 1649

RESULT 5

Q9PF38

ID Q9PF38 PRELIMINARY; PRT; 612 AA.  
 AC Q9PF38;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Beta-galactosidase (EC 3.2.1.23) (Lactase).  
 GN XF0840.  
 OS *Xylella fastidiosa*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; *Xylella*.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9a5c;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";   
 RL Nature 406:151-159(2000).  
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
 CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.  
 DR EMBL; AE003923; AAF83650.1; -.  
 DR PIR; E82756; E82756.  
 DR GO; GO:0009341; C:beta-galactosidase complex; IEA.  
 DR GO; GO:0004565; F:beta-galactosidase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR008979; Gal\_bind\_like.  
 DR InterPro; IPR001944; Glyco\_hydro\_35.  
 DR Pfam; PF01301; Glyco\_hydro\_35; 1.  
 DR PRINTS; PR00742; GLHYDRLASE35.  
 DR PROSITE; PS01182; GLYCOSYL\_HYDROL\_F35; 1.



KW Glycosidase; Hydrolase; Complete proteome.

SQ SEQUENCE 612 AA; 67893 MW; B894D6A3EB789BCB CRC64;

Query Match 92.9%; Score 39; DB 16; Length 612;

Best Local Similarity 85.7%; Pred. No. 98;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7

:|||||

Db 356 LQPPPLP 362

# RESULT 6

Q87AJ2

ID Q87AJ2 PRELIMINARY; PRT; 612 AA.

AC Q87AJ2;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Beta-galactosidase.

GN BGA OR PD1833.

OS Xylella fastidiosa (strain Temecula1 / ATCC 700964).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI\_TaxID=183190;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22421331; PubMed=12533478;

RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,

RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,

RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,

RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,

RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,

RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,

RA Marino C.L., Giglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,

RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,

RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,

RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,

RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,

RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,

RA Kitajima J.P.;

RT "Comparative analyses of the complete genome sequences of Pierce's

RT disease and citrus variegated chlorosis strains of Xylella

RT fastidiosa.";

RL J. Bacteriol. 185:1018-1026(2003).

DR EMBL; AE012560; AAO29665.1; -.

DR GO; GO:0009341; C:beta-galactosidase complex; IEA.

DR GO; GO:0004565; F:beta-galactosidase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR008979; Gal\_bind\_like.

DR InterPro; IPR001944; Glyco\_hydro\_35.

DR Pfam; PF01301; Glyco\_hydro\_35; 1.

DR PRINTS; PR00742; GLHYDRLASE35.

DR PROSITE; PS01182; GLYCOSYL\_HYDROL\_F35; 1.

KW Complete proteome.

SQ SEQUENCE 612 AA; 67992 MW; 2B66663591EB556B CRC64;

Query Match 92.9%; Score 39; DB 16; Length 612;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 356 LQPPPLP 362

RESULT 7

Q9GRR6

ID Q9GRR6 PRELIMINARY; PRT; 1404 AA.  
AC Q9GRR6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN L5882.01.  
OS Leishmania major.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Friedlin;  
RA Masuy D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,  
RA Rajandream M.A., Barrell B.G.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Friedlin;  
RX MEDLINE=98146435; PubMed=9477341;  
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
RA Smith D.F.;  
RT "A physical map of the Leishmania major Friedlin genome.";  
RL Genome Res. 8:135-145(1998).  
DR EMBL; AL445943; CAC14319.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1 1  
SQ SEQUENCE 1404 AA; 152867 MW; 25F42AB6702B9D3C CRC64;

Query Match 92.9%; Score 39; DB 5; Length 1404;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 1142 LQPPPLP 1148

RESULT 8

Q9SGY5

ID Q9SGY5 PRELIMINARY; PRT; 542 AA.  
AC Q9SGY5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE F20B24.8.

OS *Arabidopsis thaliana* (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
 RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F20B24 from chromosome  
 RT I.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES  
 CC (POLYGALACTURONASES).  
 DR EMBL; AC009398; AAF17670.1; -.  
 DR PIR; H86239; H86239.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0004650; F:polygalacturonase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR000743; Glyco\_hydro\_28.  
 DR InterPro; IPR006626; PbH1.  
 DR Pfam; PF00295; Glyco\_hydro\_28; 1.  
 DR SMART; SM00710; PbH1; 4.  
 DR PROSITE; PS00502; POLYGALACTURONASE; 1.  
 KW Cell wall; Glycosidase; Hydrolase.  
 SQ SEQUENCE 542 AA; 59705 MW; 7C67D0C3BCD19BC8 CRC64;

Query Match 90.5%; Score 38; DB 10; Length 542;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 105 VQPPPLP 111

# RESULT 9

## Q8TES7

ID Q8TES7 PRELIMINARY; PRT; 678 AA.  
 AC Q8TES7;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE FLJ00103 protein (Fragment).  
 GN FLJ00103.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Spleen;  
 RA Ohara O., Nagase T., Kikuno R., Okumura K.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 spleen.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK074045; BAB84871.1; -.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 678 AA; 70813 MW; 4847E47E56F5D696 CRC64;

Query Match 90.5%; Score 38; DB 4; Length 678;  
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 652 VQPPPLP 658

# RESULT 10

## Q8LIE4

ID Q8LIE4 PRELIMINARY; PRT; 812 AA.  
 AC Q8LIE4;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE OJ1316\_A04.20 protein (P0427D10.3 protein).  
 GN OJ1316\_A04.20 OR P0427D10.3.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC  
 clone:OJ1316\_A04.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC  
 clone:P0427D10.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AP003822; BAC07001.1; -.  
 DR EMBL; AP004272; BAC20038.1; -.  
 DR Gramene; Q8LIE4; -.  
 DR InterPro; IPR008938; ARM.  
 SQ SEQUENCE 812 AA; 87347 MW; B58C5223F0385634 CRC64;

Query Match 90.5%; Score 38; DB 10; Length 812;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7

Db                   :|||||  
26 VQPPPLP 32

RESULT 11

Q8T1F5

ID    Q8T1F5           PRELIMINARY;           PRT;   2165 AA.  
AC    Q8T1F5;  
DT    01-JUN-2002 (TrEMBLrel. 21, Created)  
DT    01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE    Similar to Dictyostelium discoideum (Slime mold). protein tyrosine  
DE    kinase.  
OS    Dictyostelium discoideum (Slime mold).  
OC    Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
OX    NCBI\_TaxID=44689;  
RN    [1]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=AX4;  
RX    MEDLINE=22092622; PubMed=12097910;  
RA    Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA    Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA    Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
RT    "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."  
RL    Nature 418:79-85(2002).  
RN    [2]  
RP    SEQUENCE FROM N.A.  
RC    STRAIN=AX4;  
RA    Baumgart C.;  
RL    Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.  
DR    EMBL; AC116959; AAM08443.2; -.  
DR    GO; GO:0005524; F:ATP binding; IEA.  
DR    GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR    GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR    GO; GO:0007242; P:intracellular signaling cascade; IEA.  
DR    GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR    InterPro; IPR001849; PH.  
DR    InterPro; IPR000719; Prot\_kinase.  
DR    InterPro; IPR001683; PX.  
DR    InterPro; IPR002290; Ser\_thr\_pkinase.  
DR    InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
DR    InterPro; IPR001245; Tyr\_pkinase.  
DR    PRINTS; PR00109; TYRKINASE.  
DR    ProDom; PD0000001; Prot\_kinase; 1.  
DR    SMART; SM00220; S\_TKc; 1.  
DR    SMART; SM00219; TyrKc; 1.  
DR    PROSITE; PS50003; PH\_DOMAIN; 1.  
DR    PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR    PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR    PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR    PROSITE; PS50195; PX; 1.  
KW    Kinase.  
SQ    SEQUENCE    2165 AA;   241910 MW;   AFFF8E5B1C32202F CRC64;

Query Match                   90.5%;   Score 38;   DB 5;   Length 2165;  
Best Local Similarity   85.7%;   Pred. No. 4.6e+02;  
Matches       6;   Conservative   1;   Mismatches    0;   Indels       0;   Gaps       0;

Qy 1 MQPPPLP 7  
:|||||  
Db 1284 IQPPPLP 1290

RESULT 12

Q8IC57

ID Q8IC57 PRELIMINARY; PRT; 2879 AA.  
AC Q8IC57;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Erythrocyte membrane protein 1 (PfEMP1).  
GN VAR.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,  
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL844505; CAD50587.1; -.  
DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.  
DR GO; GO:0009405; P:pathogenesis; IEA.  
DR InterPro; IPR004258; PFEMP.  
DR Pfam; PF03011; PFEMP; 1.  
SQ SEQUENCE 2879 AA; 333477 MW; 6A2EEACE2C92C6C1 CRC64;

Query Match 90.5%; Score 38; DB 5; Length 2879;  
Best Local Similarity 85.7%; Pred. No. 6e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
:|||||  
Db 2429 VQPPPLP 2435

RESULT 13

Q85G37

ID Q85G37 PRELIMINARY; PRT; 44 AA.  
AC Q85G37;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ORF44.  
OS Cyanidioschyzon merolae (Red alga).  
OG Chloroplast.  
OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.  
OX NCBI\_TaxID=45157;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=10D;  
RX MEDLINE=22639682; PubMed=12755171;  
RA Ohta N., Matsuzaki M., Misumi O., Miyagishima S., Nozaki H.,

RA Tanaka K., Shin-i T., Kohara Y., Kuroiwa T.;  
 RT "Complete Sequence and Analysis of the Plastid Genome of the  
 RT Unicellular Red Alga *Cyanidioschyzon merolae*.";  
 RL DNA Res. 10:67-77(2003).  
 DR EMBL; AB002583; BAC76154.1; -.  
 DR GO; GO:0009507; C:chloroplast; IEA.  
 KW Chloroplast.  
 SQ SEQUENCE 44 AA; 4903 MW; A38CAD0D007AA7B0 CRC64;

Query Match 88.1%; Score 37; DB 8; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 27 QPPPLP 32

# RESULT 14

Q87610

ID Q87610 PRELIMINARY; PRT; 72 AA.  
 AC Q87610;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Rev protein (Fragment).  
 GN REV.  
 OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11723;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=P031;  
 RX MEDLINE=97138325; PubMed=8985351;  
 RA Bibollet-Ruche F., Brengues C., Galat-Luong A., Galat G., Pourrut X.,  
 RA Vidal N., Veas F., Durand J.P., Cuny G.;  
 RT "Genetic diversity of simian immunodeficiency viruses from West  
 RT African green monkeys: evidence of multiple genotypes within  
 RT populations from the same geographical locale.";  
 RL J. Virol. 71:307-313(1997).  
 DR EMBL; U37213; AAC56168.1; -.  
 DR GO; GO:0042025; C:host cell nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000625; REV\_protein.  
 DR Pfam; PF00424; REV; 1.  
 FT NON\_TER 1 1  
 FT NON\_TER 72 72  
 SQ SEQUENCE 72 AA; 8188 MW; B7A8027FA9BA432A CRC64;

Query Match 88.1%; Score 37; DB 15; Length 72;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 58 QPPPLP 63

RESULT 15

Q87603

ID Q87603 PRELIMINARY; PRT; 73 AA.  
AC Q87603;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Rev protein (Fragment).  
GN REV.  
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11723;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=P058;  
RX MEDLINE=97138325; PubMed=8985351;  
RA Bibollet-Ruche F., Brengues C., Galat-Luong A., Galat G., Pourrut X.,  
RA Vidal N., Veas F., Durand J.P., Cuny G.;  
RT "Genetic diversity of simian immunodeficiency viruses from West  
RT African green monkeys: evidence of multiple genotypes within  
RT populations from the same geographical locale."  
RL J. Virol. 71:307-313(1997).  
DR EMBL; U37210; AAC56159.1; -.  
DR GO; GO:0042025; C:host cell nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR000625; REV\_protein.  
DR Pfam; PF00424; REV; 1.  
FT NON\_TER 1 1  
SQ SEQUENCE 73 AA; 8217 MW; A31FDD937DC9BB4D CRC64;

Query Match 88.1%; Score 37; DB 15; Length 73;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||||  
Db 58 QPPPLP 63

Search completed: August 24, 2004, 15:50:24  
Job time : 27.6269 secs



OM protein - protein search, using sw model

Run on: August 24, 2004, 14:57:04 ; Search time 3.76119 Seconds  
 (without alignments)  
 96.908 Million cell updates/sec

Title: US-09-641-801-1  
 Perfect score: 42  
 Sequence: 1 MQPPPLP 7

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	39	92.9	222	1	COLI_THUOB	Q9y9k2 t corticotr
2	39	92.9	701	1	CG1_HUMAN	Q13495 homo sapien
3	39	92.9	1074	1	EMI2_MOUSE	Q8k482 mus musculu
4	38	90.5	536	1	ARP_ARATH	P45951 arabidopsis
5	37	88.1	167	1	SERO_GALME	O76192 galleria me
6	37	88.1	211	1	RGM1_YEAST	Q00453 saccharomyc
7	37	88.1	278	1	TNF6_RAT	P36940 rattus norv
8	37	88.1	330	1	PEXD_CAEEL	Q19951 caenorhabdi
9	37	88.1	393	1	WR25_ARATH	O22921 arabidopsis
10	37	88.1	483	1	B1AR_MELGA	P07700 meleagris g
11	37	88.1	562	1	ILVD_PYRAE	Q8zyu6 pyrobaculum
12	37	88.1	620	1	EXTN_TOBAC	P13983 nicotiana t
13	37	88.1	649	1	DREB_HUMAN	Q16643 homo sapien
14	37	88.1	745	1	AXN_DROME	Q9v407 drosophila
15	37	88.1	830	1	DYN1_CAEEL	P39055 caenorhabdi
16	37	88.1	1027	1	DOC1_MOUSE	Q8bur4 mus musculu
17	37	88.1	1032	1	S24D_HUMAN	O94855 homo sapien

18	37	88.1	1124	1	TCF8_HUMAN	P37275	homo sapien
19	37	88.1	1567	1	FMN2_MOUSE	Q9jl04	mus musculu
20	37	88.1	3312	1	CLR3_HUMAN	Q9nyq7	homo sapien
21	36	85.7	463	1	YAD4_YEAST	P28003	saccharomyc
22	36	85.7	505	1	WASL_BOVIN	Q95107	bos taurus
23	36	85.7	629	1	RA21_XENLA	O93310	xenopus lae
24	36	85.7	736	1	DVL2_XENLA	P51142	xenopus lae
25	36	85.7	823	1	NUC1_NEUCR	P20824	neurospora
26	36	85.7	1484	1	CES2_HUMAN	Q9bxf3	homo sapien
27	35	83.3	175	1	OE56_NPVHZ	O10620	heliothis z
28	35	83.3	414	1	PRS6_CAEEL	P46502	caenorhabdi
29	35	83.3	533	1	MUTL_THEAQ	P96082	thermus aqu
30	35	83.3	587	1	ACA2_YEAST	P40535	saccharomyc
31	35	83.3	588	1	DAB1_MOUSE	P97318	mus musculu
32	35	83.3	639	1	SF01_HUMAN	Q15637	homo sapien
33	35	83.3	653	1	SF01_MOUSE	Q64213	mus musculu
34	35	83.3	666	1	BCHD_HELMO	Q9zge6	heliobacill
35	35	83.3	808	1	Y066_NPVAC	P41467	autographa
36	35	83.3	1271	1	RBMG_HUMAN	Q9upn6	homo sapien
37	35	83.3	2476	1	ATRX_MOUSE	Q61687	mus musculu
38	35	83.3	2492	1	ATRX_HUMAN	P46100	homo sapien
39	35	83.3	2492	1	ATRX_PANTR	Q7yqm4	pan troglod
40	35	83.3	2492	1	ATRX_PONPY	Q7yqm3	pongo pygma
41	34	81.0	91	1	Y879_MYCTU	Q10541	mycobacteri
42	34	81.0	193	1	PGC1_PIG	Q95250	sus scrofa
43	34	81.0	194	1	PGC1_HUMAN	O00264	homo sapien
44	34	81.0	194	1	PGC1_MOUSE	O55022	mus musculu
45	34	81.0	194	1	PGC1_RAT	P70580	rattus norv

# ALIGNMENTS

## RESULT 1

COLI\_THUOB

ID COLI\_THUOB STANDARD; PRT; 222 AA.

AC Q9YGK2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Corticotropin-lipotropin precursor (Pro-opiomelanocortin) (POMC)

DE [Contains: Corticotropin (Adrenocorticotropic hormone) (ACTH);

DE Melanotropin alpha (Alpha-MSH); Corticotropin-like intermediary

DE peptide (CLIP); Lipotropin beta (Beta-LPH); Lipotropin gamma (Gamma-

DE LPH); Melanotropin beta (Beta-MSH); Beta-endorphin; Met-enkephalin].

OS Thunnus obesus (Bigeye tuna).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;

OC Scombridae; Thunnus.

OX NCBI\_TaxID=8241;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Pituitary;

RA Amemiya Y., Takahashi A., Kawauchi H.;

RT "Tuna proopiomelanocortin cDNA.";

RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.

CC -!- PTM: Specific enzymatic cleavages at paired basic residues yield  
 CC the different active peptides.  
 CC -!- SIMILARITY: Belongs to the POMC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB020971; BAA35125.1; -.  
 DR InterPro; IPR001941; Mcortin\_ACTH.  
 DR Pfam; PF00976; ACTH\_domain; 1.  
 DR PRINTS; PR00383; MELANOCORTIN.  
 DR ProDom; PD003250; Mcortin\_ACTH; 1.  
 KW Endorphin; Hormone; Cleavage on pair of basic residues; Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PEPTIDE 93 132 CORTICOTROPIN.  
 FT PEPTIDE 93 107 MELANOTROPIN ALPHA.  
 FT PEPTIDE 111 132 CORTICOTROPIN-LIKE INTERMEDIARY PEPTIDE.  
 FT PEPTIDE 136 222 LIPOTROPIN BETA.  
 FT PEPTIDE 136 188 LIPOTROPIN GAMMA.  
 FT PEPTIDE 172 188 MELANOTROPIN BETA.  
 FT PEPTIDE 191 222 BETA-ENDORPHIN.  
 FT PEPTIDE 191 195 MET-ENKEPHALIN.  
 FT DOMAIN 71 75 POLY-SER.  
 FT DOMAIN 80 87 POLY-SER.  
 SQ SEQUENCE 222 AA; 24970 MW; 7868C713D6360B70 CRC64;

Query Match 92.9%; Score 39; DB 1; Length 222;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 62 LQPPPLP 68

## RESULT 2

### CG1\_HUMAN

ID CG1\_HUMAN STANDARD; PRT; 701 AA.  
 AC Q13495;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE CG1 protein (F18).  
 GN CXORF6 OR CG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96225444; PubMed=8640223;  
 RA Laporte J., Hu L.-J., Kretz C., Mandel J.-L., Kioschis P., Coy J.,

RA Klauck S.M., Poutska A., Dahl N.;  
 RT "A gene mutated in X-linked myotubular myopathy defines a new  
 RT putative tyrosine phosphatase family conserved in yeast."  
 RL Nat. Genet. 13:175-182(1996).  
 CC -!- TISSUE SPECIFICITY: Preferentially expressed in skeletal muscle.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U46023; AAC50551.1; -.  
 DR Genew; HGNC:2568; CXorf6.  
 DR MIM; 300120; -.  
 FT DOMAIN 275 286 POLY-PRO.  
 FT DOMAIN 419 429 POLY-GLN.  
 FT DOMAIN 522 533 POLY-GLN.  
 SQ SEQUENCE 701 AA; 74475 MW; AD1C18E4F6DFD34E CRC64;

Query Match 92.9%; Score 39; DB 1; Length 701;  
 Best Local Similarity 85.7%; Pred. No. 77;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 273 LQPPPLP 279

# RESULT 3

## EMI2\_MOUSE

ID EMI2\_MOUSE STANDARD; PRT; 1074 AA.  
 AC Q8K482;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE EMILIN 2 precursor (Elastin microfibril interface-located protein 2)  
 DE (Elastin microfibril interfacier 2) (Basilin).  
 GN EMILIN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=C57BL/6J; TISSUE=Cochlea;  
 RX MEDLINE=22722101; PubMed=12837629;  
 RA Amma L.L., Goodyear R., Faris J.S., Jones I., Ng L., Richardson G.,  
 RA Forrest D.;  
 RT "An emilin family extracellular matrix protein identified in the  
 RT cochlear basilar membrane."  
 RL Mol. Cell. Neurosci. 23:460-472(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Limb;

RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=22209197; PubMed=12221002;  
 RA Leimeister C., Steidl C., Schumacher N., Erhard S., Gessler M.;  
 RT "Developmental expression and biochemical characterization of Emu  
 RT family members.";  
 RL Dev. Biol. 249:204-218(2002).  
 CC -!- FUNCTION: May be responsible for anchoring smooth muscle cells to  
 CC elastic fibers, and may be involved not only in the formation of  
 CC the elastic fiber, but also in the processes that regulate vessel  
 CC assembly. Has cell adhesive capacity. Major component of the  
 CC cochlear basilar membrane (BM) which may contribute to the  
 CC developmental assembly or function of the BM.  
 CC -!- SUBUNIT: Homotrimer associated through a moderately stable  
 CC interaction of the C-terminal globular Clq domains, allowing the  
 CC nucleation of the triple helix and then a further quaternary  
 CC assembly to higher order polymers via intermolecular disulfide  
 CC bonds (By similarity). Interacts with EMILIN1.  
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix. Found mainly at the  
 CC interface between amorphous elastin and microfibrils.  
 CC -!- TISSUE SPECIFICITY: Highest levels are present in cochlea of P8  
 CC pups, followed by modest levels in adult heart and lung, and much  
 CC lower levels in forebrain, brainstem, cerebellum and hypothalamus.  
 CC Very low levels detected in muscle, liver, kidney and eye.  
 CC -!- DEVELOPMENTAL STAGE: Low levels detected in cochlea in neonatal  
 CC pups at P1. Levels increased 2-fold by P5 and rose further to 16-  
 CC fold at P13. Expression declined somewhat in adult mice. At E9.5,  
 CC as during all stages of development, it is strongly expressed in  
 CC the neural fold, the limb buds and the heart.  
 CC -!- SIMILARITY: Contains 1 EMI domain.  
 CC -!- SIMILARITY: Contains 1 collagenous domain.  
 CC -!- SIMILARITY: Contains 1 ClQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF468645; AAM53532.1; -.  
 DR EMBL; BC053753; AAH53753.1; -.  
 DR MGD; MGI:2389136; Emilin2.  
 DR InterPro; IPR001073; Clq.  
 DR InterPro; IPR008983; TNF\_like.  
 DR Pfam; PF00386; Clq; 1.  
 DR SMART; SM00110; ClQ; 1.  
 KW Cell adhesion; Extracellular matrix; Signal; Coiled coil; Collagen;  
 KW Glycoprotein.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT CHAIN 34 1074 EMILIN 2.  
 FT DOMAIN 48 121 EMI.  
 FT DOMAIN 181 218 COILED COIL (POTENTIAL).  
 FT DOMAIN 259 345 COILED COIL (POTENTIAL).  
 FT DOMAIN 374 394 COILED COIL (POTENTIAL).  
 FT DOMAIN 533 554 COILED COIL (POTENTIAL).  
 FT DOMAIN 582 620 COILED COIL (POTENTIAL).  
 FT DOMAIN 783 851 PRO-RICH.  
 FT DOMAIN 852 913 COLLAGEN-LIKE.  
 FT DOMAIN 914 924 PRO-RICH.  
 FT DOMAIN 928 1067 ClQ.  
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 744 744 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 995 995 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1074 AA; 117310 MW; 4B81728C83CE52A7 CRC64;

Query Match 92.9%; Score 39; DB 1; Length 1074;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQPPPLP 7  
 :|||||  
 Db 842 LQPPPLP 848

#### RESULT 4

##### ARP\_ARATH

ID ARP\_ARATH STANDARD; PRT; 536 AA.  
 AC P45951;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Apurinic endonuclease-redox protein (DNA-(apurinic or apyrimidinic  
 DE site) lyase) (EC 4.2.99.18).  
 GN ARP OR REF OR AT2G41460 OR T26J13.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE OF 10-536 FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Callus;  
 RX MEDLINE=94211851; PubMed=7512729;  
 RA Babiychuk E., Kushnir S., van Montagu M., Inze D.;  
 RT "The Arabidopsis thaliana apurinic endonuclease Arp reduces human  
 RT transcription factors Fos and Jun.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3299-3303(1994).  
 CC -!- FUNCTION: Repairs oxidative DNA damages, seems also to act as a  
 CC redox factor. Is multifunctional and may be involved both in DNA  
 CC repair and in the regulation of transcription.  
 CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
 CC apyrimidinic site in DNA is broken by a beta-elimination reaction,  
 CC leaving a 3'-terminal unsaturated sugar and a product with a  
 CC terminal 5'-phosphate.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Expressed in the siliques, flowers, and stems.  
 CC A high level expression is seen in the leaves.  
 CC -!- SIMILARITY: Belongs to the DNA repair enzymes AP/exoA family.  
 CC -!- SIMILARITY: Contains 1 SAP domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AC004625; AAC23731.1; -.  
 DR EMBL; X76912; CAA54234.1; -.  
 DR PIR; T02441; T02441.  
 DR HSSP; P27695; 1E9N.  
 DR InterPro; IPR000097; APendonclsel.  
 DR InterPro; IPR005135; Exo\_endo\_phos.  
 DR InterPro; IPR004808; ExoIII\_xth.  
 DR InterPro; IPR003034; SAP.  
 DR Pfam; PF03372; Exo\_endo\_phos; 1.  
 DR Pfam; PF02037; SAP; 1.  
 DR SMART; SM00513; SAP; 1.

DR TIGRFAMs; TIGR00195; exoDNase\_III; 1.  
 DR TIGRFAMs; TIGR00633; xth; 1.  
 DR PROSITE; PS00726; AP\_NUCLEASE\_F1\_1; 1.  
 DR PROSITE; PS00727; AP\_NUCLEASE\_F1\_2; 1.  
 DR PROSITE; PS00728; AP\_NUCLEASE\_F1\_3; 1.  
 DR PROSITE; PS50800; SAP; 1.  
 KW DNA repair; Lyase; Nuclear protein.  
 FT DOMAIN 1 278 HIGHLY CHARGED; INCREASES THE AFFINITY OF  
 FT ARP FOR DNA.  
 FT DOMAIN 97 131 SAP.  
 FT DOMAIN 279 536 AP ENDONUCLEASE.  
 FT METAL 313 313 MAGNESIUM OR MANGANESE (BY SIMILARITY).  
 FT ACT\_SITE 527 527 GENERAL BASE (BY SIMILARITY).  
 SQ SEQUENCE 536 AA; 60260 MW; 5C1FC17EA991D27B CRC64;

Query Match 90.5%; Score 38; DB 1; Length 536;  
 Best Local Similarity 85.7%; Pred. No. 81;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQPPPLP 7  
 |:|||||  
 Db 264 MRPPPLP 270

# RESULT 5

## SERO\_GALME

ID SERO\_GALME STANDARD; PRT; 167 AA.  
 AC 076192;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Seroin precursor (Silk 23 kDa glycoprotein).  
 OS Galleria mellonella (Wax moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Pyraloidea;  
 OC Pyralidae; Galleriinae; Galleria.  
 OX NCBI\_TaxID=7137;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-31.  
 RC TISSUE=Silk gland;  
 RX MEDLINE=98288272; PubMed=9624126;  
 RA Zurovec M., Yang C., Kodrik D., Sehnal F.;  
 RT "Identification of a novel type of silk protein and regulation of its  
 RT expression."  
 RL J. Biol. Chem. 273:15423-15428(1998).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: PRODUCED BY BOTH THE POSTERIOR (PSG) AND  
 CC MIDDLE (MSG) SECTIONS OF SILK GLANDS.  
 CC -!- DEVELOPMENTAL STAGE: Seroin mRNA is high in the silk glands of  
 CC feeding larvae, declines at ecdysis, reaches a maximum during  
 CC cocoon spinning, and thereafter rapidly drops to an undetectable  
 CC level.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF009828; AAC25171.1; -.  
KW Silk; Glycoprotein; Signal; Repeat.  
FT SIGNAL 1 17  
FT CHAIN 18 167 SEROIN.  
FT REPEAT 38 46 1-1.  
FT REPEAT 56 64 1-2.  
FT REPEAT 76 78 2-1.  
FT REPEAT 79 81 2-2.  
FT REPEAT 82 84 2-3.  
FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 167 AA; 18088 MW; 27A6ABE862774EB9 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 167;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 QPPPLP 7  
| | | | |  
Db 42 QPPPLP 47

#### RESULT 6

##### RGM1\_YEAST

ID RGM1 YEAST STANDARD; PRT; 211 AA.  
AC Q00453;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable transcription repressor protein RGM1.  
GN RGM1 OR YMR182C OR YM8010.12C.  
OS *Saccharomyces cerevisiae* (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=92020118; PubMed=1923755;  
RA Estruch F.;  
RT "The yeast putative transcriptional repressor RGM1 is a proline-rich  
RT zinc finger protein.";  
RL Nucleic Acids Res. 19:4873-4877(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=97313268; PubMed=9169872;  
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
RT XIII.";

RL Nature 387:90-93(1997).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- INDUCTION: UNDER THE CONTROL OF THE INDUCIBLE GAL10 PROMOTER.  
 CC -!- DOMAIN: THE PRO-RICH REGION OF RGM1 ATTACHED TO A HETEROLOGOUS DNA  
 CC BINDING DOMAIN IS ABLE TO REPRESS THE EXPRESSION OF THE TARGET  
 CC GENE.  
 CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59861; CAA42521.1; -.  
 DR EMBL; Z49808; CAA89915.1; -.  
 DR PIR; S55129; S55129.  
 DR HSSP; P07248; 2ADR.  
 DR GermOnline; 142856; -.  
 DR TRANSFAC; T00730; -.  
 DR SGD; S0004794; RGM1.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; IDA.  
 DR GO; GO:0000122; P:negative regulation of transcription from P. . .; IDA.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 2.  
 DR SMART; SM00355; ZnF\_C2H2; 2.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
 KW Transcription regulation; Repressor; DNA-binding; Zinc-finger;  
 KW Metal-binding; Nuclear protein; Repeat.  
 FT DOMAIN 6 11 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT ZN\_FING 19 44 C2H2-TYPE 1.  
 FT ZN\_FING 50 73 C2H2-TYPE 2.  
 FT DOMAIN 95 211 PRO-RICH.  
 FT CONFLICT 114 114 A -> V (IN REF. 1).  
 SQ SEQUENCE 211 AA; 23855 MW; FADCE4757FE7D372 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 154 QPPPLP 159

RESULT 7  
 TNF6\_RAT  
 ID TNF6\_RAT STANDARD; PRT; 278 AA.  
 AC P36940;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Tumor necrosis factor ligand superfamily member 6 (FAS antigen)

DE ligand).  
 GN TNFSF6 OR FASL OR APT1LG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94084792; PubMed=7505205;  
 RA Suda T., Takahashi T., Golstein P., Nagata S.;  
 RT "Molecular cloning and expression of the Fas ligand, a novel member  
 RT of the tumor necrosis factor family.";  
 RL Cell 75:1169-1178(1993).  
 CC -!- FUNCTION: Cytokine that binds to TNFRSF6/FAS, a receptor that  
 CC transduces the apoptotic signal into cells. May be involved in  
 CC cytotoxic T cell mediated apoptosis and in T cell development.  
 CC TNFRSF6/FAS-mediated apoptosis may have a role in the induction of  
 CC peripheral tolerance, in the antigen-stimulated suicide of mature  
 CC T cells, or both. Binding to the decoy receptor TNFRSF6B/DcR3  
 CC modulates its effects (By similarity).  
 CC -!- SUBUNIT: Homotrimer (Probable).  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein and secreted (By  
 CC similarity).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ACTIVATED SPLENOCYTES AND  
 CC THYMOCYTES. MODERATE OR WEAK EXPRESSION FOUND IN SMALL INTESTINES,  
 CC KIDNEY AND LUNG.  
 CC -!- INDUCTION: BY PMA/IONOMYCIN AND CONCAVALIN/INTERLEUKIN-2.  
 CC -!- PTM: The soluble form derives from the membrane form by  
 CC proteolytic processing (By similarity).  
 CC -!- SIMILARITY: Belongs to the tumor necrosis factor family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U03470; AAC52129.1; -.  
 DR PIR; A49266; A49266.  
 DR HSSP; P01375; 4TSV.  
 DR InterPro; IPR008064; Fas\_ligand.  
 DR InterPro; IPR006053; TNF\_abc.  
 DR InterPro; IPR006052; TNF\_family.  
 DR InterPro; IPR008983; TNF\_like.  
 DR InterPro; IPR003636; TNF\_subf.  
 DR Pfam; PF00229; TNF; 1.  
 DR PRINTS; PR01681; FASLIGAND.  
 DR PRINTS; PR01234; TNECROSISFCT.  
 DR ProDom; PD002012; TNF\_subf; 1.  
 DR SMART; SM00207; TNF; 1.  
 DR PROSITE; PS00251; TNF\_1; 1.  
 DR PROSITE; PS50049; TNF\_2; 1.  
 KW Cytokine; Apoptosis; Transmembrane; Glycoprotein; Signal-anchor.  
 FT CHAIN 1 278 TUMOR NECROSIS FACTOR LIGAND SUPERFAMILY  
 FT MEMBER 6, MEMBRANE FORM.

FT	CHAIN	127	278	TUMOR NECROSIS FACTOR LIGAND SUPERFAMILY
FT				MEMBER 6, SOLUBLE FORM (BY SIMILARITY).
FT	DOMAIN	1	77	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	78	99	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT				(POTENTIAL).
FT	DOMAIN	100	278	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	4	69	PRO-RICH.
FT	DOMAIN	45	58	POLY-PRO.
FT	SITE	126	127	CLEAVAGE (BY SIMILARITY).
FT	DISULFID	199	230	POTENTIAL.
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	247	247	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	257	257	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	278 AA; 31140 MW; 2898E18A862CEAC6 CRC64;		

Query Match 88.1%; Score 37; DB 1; Length 278;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 60 QPPPLP 65

# RESULT 8

## PEXD\_CAEEL

ID PEXD CAEEL STANDARD; PRT; 330 AA.  
 AC Q19951;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable peroxisomal membrane protein PEX13 (Peroxin-13).  
 GN F32A5.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Pauley A.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Component of the peroxisomal translocation machinery  
 CC with PEX14 and PEX17. Functions as a docking factor for the  
 CC predominantly cytoplasmic PTS1 receptor (PAS10/PEX5) (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U20864; AAC46665.2; -.  
DR HSSP; P00519; 1ABL.  
DR WormPep; F32A5.6; CE28290.  
DR InterPro; IPR007223; Peroxin-13\_N.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF04088; Peroxin-13\_N; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW Peroxisome; Transport; Protein transport; Transmembrane; SH3 domain.  
FT DOMAIN 1 192 LUMENAL (POTENTIAL).  
FT TRANSMEM 193 213 POTENTIAL.  
FT DOMAIN 214 330 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 235 300 SH3.  
SQ SEQUENCE 330 AA; 35765 MW; 7DDD647F07A04403 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 330;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
| | | | |  
Db 8 QPPPLP 13

#### RESULT 9

##### WR25\_ARATH

ID WR25\_ARATH STANDARD; PRT; 393 AA.  
AC O22921; Q94AT4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable WRKY transcription factor 25 (WRKY DNA-binding protein 25).  
GN WRKY25 OR AT2G30250 OR T9D9.6.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=Flower;  
RA Ulker B., Kushnir S., Somssich I.E.;  
RT "Arabidopsis thaliana transcription factor WRKY25.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,  
 RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,  
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,  
 RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,  
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,  
 RA Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*  
 RT *thaliana*.";  
 RL Nature 402:761-768(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Empirical analysis of transcriptional activity in the *Arabidopsis*  
 RT genome.";  
 RL Science 302:842-846(2003).  
 CC -!- FUNCTION: Transcription factor. Interacts specifically with the W  
 CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-  
 CC responsive cis-acting element (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: Belongs to the WRKY group I family.  
 CC -!- SIMILARITY: Contains 2 WRKY domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF418309; AAL13040.1; -.  
 DR EMBL; AC002338; AAC16930.1; -.  
 DR EMBL; AC004165; AAM14918.1; -.  
 DR EMBL; AY062720; AAL32798.1; -.  
 DR EMBL; AY114650; AAM47969.1; -.  
 DR EMBL; AY136318; AAM96984.1; -.  
 DR EMBL; AY045813; AAK76487.2; -.  
 DR EMBL; BT002338; AAN86171.1; -.  
 DR PIR; T00575; T00575.  
 DR InterPro; IPR003657; WRKY.  
 DR Pfam; PF03106; WRKY; 2.  
 DR PROSITE; PS50811; WRKY; 2.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.

FT DNA\_BIND 160 224 WRKY 1.  
 FT DNA\_BIND 322 387 WRKY 2.  
 SQ SEQUENCE 393 AA; 44133 MW; 9057D65B69E12A0C CRC64;

Query Match 88.1%; Score 37; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 50 QPPPLP 55

# RESULT 10

## BLAR\_MELGA

ID BLAR\_MELGA STANDARD; PRT; 483 AA.  
 AC P07700;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Beta-1 adrenergic receptor (Beta-T).  
 OS Meleagris gallopavo (Common turkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.  
 OX NCBI\_TaxID=9103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86313664; PubMed=3018746;  
 RA Yarden Y., Rodriguez H., Wong S.K.-F., Brandt D.R., May D.C.,  
 RA Burnier J., Harkins R.N., Chen E.Y., Ramachandran J., Ullrich A.,  
 RA Ross E.M.;  
 RT "The avian beta-adrenergic receptor: primary structure and membrane  
 RT topology.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6795-6799(1986).  
 RN [2]  
 RP STRUCTURE BY NMR OF 345-359.  
 RX MEDLINE=95129696; PubMed=7828722;  
 RA Jung H., Windhaber R., Palm D., Schnackerz K.D.;  
 RT "NMR and circular dichroism studies of synthetic peptides derived  
 RT from the third intracellular loop of the beta-adrenoceptor.";  
 RL FEBS Lett. 358:133-136(1995).  
 CC -!- FUNCTION: BETA-ADRENERGIC RECEPTORS MEDIATE THE CATECHOLAMINE-  
 CC INDUCED ACTIVATION OF ADENYLATE CYCLASE THROUGH THE ACTION OF G  
 CC PROTEINS. THIS RECEPTOR BINDS EPINEPHRINE AND NOREPINEPHRINE WITH  
 CC APPROXIMATIVELY EQUAL AFFINITY.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- PTM: Homologous desensitization of the receptor is mediated by its  
 CC phosphorylation by beta-adrenergic receptor kinase.  
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; M14379; AAA49627.1; -.
DR PIR; A25896; A25896.
DR PDB; 1DEP; 14-OCT-96.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRRHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Phosphorylation; Lipoprotein; Palmitate;
KW 3D-structure.
FT DOMAIN 1 43 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 44 67 1 (POTENTIAL).
FT DOMAIN 68 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 94 2 (POTENTIAL).
FT DOMAIN 95 115 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 116 137 3 (POTENTIAL).
FT DOMAIN 138 159 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 160 189 4 (POTENTIAL).
FT DOMAIN 190 205 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 206 228 5 (POTENTIAL).
FT DOMAIN 229 290 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 291 315 6 (POTENTIAL).
FT DOMAIN 316 321 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 322 344 7 (POTENTIAL).
FT DOMAIN 345 483 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (PROBABLE).
FT DISULFID 114 192 BY SIMILARITY.
FT LIPID 358 358 S-palmitoyl cysteine (By similarity).
FT HELIX 347 358
SQ SEQUENCE 483 AA; 54078 MW; B11A7E71F6CCE3E4 CRC64;

```

```

Query Match 88.1%; Score 37; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 QPPPLP 7
   |||||
Db 256 QPPPLP 261

```

```

RESULT 11
ILVD_PYRAE
ID ILVD_PYRAE STANDARD; PRT; 562 AA.
AC Q8ZYU6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD).
GN ILVD OR PAE0615.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.

```



RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum*  
 RT *aerophilum*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 CC -!- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-  
 CC oxobutanoate + H(2)O.  
 CC -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).  
 CC -!- PATHWAY: Valine and isoleucine biosynthesis; fourth step.  
 CC -!- SIMILARITY: Belongs to the ilvD / edd family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE009775; AAL62897.1; -.  
 DR HAMAP; MF\_00012; -, 1.  
 DR InterPro; IPR004404; IlvD.  
 DR InterPro; IPR000581; ILVD\_EDD\_family.  
 DR Pfam; PF00920; ILVD\_EDD; 1.  
 DR ProDom; PD002691; ILVD\_EDD\_family; 1.  
 DR TIGRFAMs; TIGR00110; ilvD; 1.  
 DR PROSITE; PS00886; ILVD\_EDD\_1; 1.  
 DR PROSITE; PS00887; ILVD\_EDD\_2; 1.  
 KW Branched-chain amino acid biosynthesis; Lyase; Iron; Iron-sulfur;  
 KW 4Fe-4S; Complete proteome.  
 FT METAL 124 124 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 FT METAL 196 196 IRON-SULFUR (4FE-4S) (POTENTIAL).  
 SQ SEQUENCE 562 AA; 59751 MW; 1D4DF358B14FC866 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 562;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 358 QPPPLP 363

RESULT 12  
 EXTN\_TOBAC  
 ID EXTN\_TOBAC STANDARD; PRT; 620 AA.  
 AC P13983;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).  
 GN HRGPNT3.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamiids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi; TISSUE=Leaf;  
 RX MEDLINE=90128263; PubMed=2612909;  
 RA Keller B., Lamb C.J.;  
 RT "Specific expression of a novel cell wall hydroxyproline-rich  
 RT glycoprotein gene in lateral root initiation.";  
 RL Genes Dev. 3:1639-1646(1989).  
 CC -!- FUNCTION: Has a specialized structural function, possibly in  
 CC the mechanical penetration of the cortex and epidermis of the  
 CC main root.  
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.  
 CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
 CC GLYCOSYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13885; CAA32090.1; -.  
 DR PIR; S06733; S06733.  
 DR InterPro; IPR000480; Glutelin.  
 DR PRINTS; PR00211; GLUTELIN.  
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 KW Hydroxylation.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 620 EXTENSIN.  
 FT REPEAT 70 73 H-A-P-P.  
 FT REPEAT 148 151 H-A-P-P.  
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
 FT REPEAT 229 235 1.  
 FT REPEAT 236 242 2.  
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.  
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.  
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 620;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 425 QPPPLP 430

RESULT 13  
 DREB\_HUMAN  
 ID DREB\_HUMAN STANDARD; PRT; 649 AA.  
 AC Q16643;  
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Drebrin (Developmentally regulated brain protein).  
 GN DBN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Osteoblast;  
 RA Fisher L.W., McBride O.W., Filpula D., Ibaraki K., Young M.F.;  
 RT "Human drebrin: cDNA sequence, mRNA tissue distribution and  
 RT chromosomal localization.";  
 RL Neurosci. Res. Commun. 14:35-42(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=94030036; PubMed=8216329;  
 RA Toda M., Shirao T., Minoshima S., Shimizu N., Toya S., Uyemura K.;  
 RT "Molecular cloning of cDNA encoding human drebrin E and chromosomal  
 RT mapping of its gene.";  
 RL Biochem. Biophys. Res. Commun. 196:468-472(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Eye, and Muscle;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Drebrins might play some role in cell migration,  
 CC extension of neuronal processes and plasticity of dendrites,  
 CC respectively. Binds F-actin.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: Brain neurons. Also found in the heart,  
 CC placenta, skeletal muscle, kidney and pancreas.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U00802; AAA16256.1; -.  
DR EMBL; D17530; BAA04480.1; -.  
DR EMBL; BC000283; AAH00283.1; -.  
DR EMBL; BC007281; AAH07281.1; -.  
DR EMBL; BC007567; AAH07567.1; -.  
DR PIR; JN0809; JN0809.  
DR Genew; HGNC:2695; DBN1.  
DR MIM; 126660; -.  
DR GO; GO:0015629; C:actin cytoskeleton; TAS.  
DR GO; GO:0003779; F:actin binding; TAS.  
DR InterPro; IPR002108; Actbind\_cofln.  
DR Pfam; PF00241; cofilin\_ADF; 1.  
DR SMART; SM00102; ADF; 1.  
KW Actin-binding; Neurone.  
SQ SEQUENCE 649 AA; 71425 MW; A1730E7C5FC32D88 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 649;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
|||  
Db 363 QPPPLP 368

#### RESULT 14

##### AXN\_DROME

ID AXN\_DROME STANDARD; PRT; 745 AA.  
AC Q9V407; Q9XYC1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Axin (Axis inhibition protein) (dAxin) (d-Axin).  
GN AXN OR CG7926.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=99174088; PubMed=10073940;  
RA Hamada F., Tomoyasu Y., Takatsu Y., Nakamura M., Nagai S.-I.,  
RA Suzuki A., Fujita F., Shibuya H., Toyoshima K., Ueno N., Akiyama T.;  
RT "Negative regulation of Wingless signaling by D-axin, a Drosophila  
RT homolog of axin.";  
RL Science 283:1739-1742(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ruel L., Anthopoulos N., Goncalves J., Manoukian A.S., Woodgett J.R.;  
RT "A Drosophila homolog of the axin gene is involved in the transduction

RT of the wingless signal regulating the stability of the armadillo  
 RT protein.";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=99387984; PubMed=10457025;  
 RA Willert K., Logan C.Y., Arora A., Fish M., Nusse R.;  
 RT "A *Drosophila* Axin homolog, Daxin, inhibits Wnt signaling.";  
 RL Development 126:4165-4173(1999).  
 CC -!- FUNCTION: Inhibitor of the WG signaling pathway. Down-regulates  
 CC beta-catenin (armadillo=ARM). Probably facilitate the  
 CC phosphorylation of beta-catenin and APC by GSK3B (zeste-white  
 CC 3=ZW3).  
 CC -!- SUBUNIT: Interacts with ZW3 and ARM. The interaction between AXN  
 CC and ARM occurs via the armadillo repeats contained in ARM.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- DEVELOPMENTAL STAGE: Ubiquitously expressed throughout the  
 CC development.  
 CC -!- SIMILARITY: Contains 1 RGS domain.  
 CC -!- SIMILARITY: Contains 1 DIX domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF086811; AAD24886.1; -.  
 DR EMBL; AF091813; AAF21293.1; -.  
 DR EMBL; AE003772; AAF56993.1; -.  
 DR FlyBase; FBgn0026597; Axn.  
 DR GO; GO:0007455; P:eye-antennal disc metamorphosis; IMP.  
 DR GO; GO:0016055; P:Wnt receptor signaling pathway; IGI.  
 DR InterPro; IPR001158; DIX.  
 DR InterPro; IPR000342; Regl\_Gprotein.  
 DR Pfam; PF00778; DIX; 1.  
 DR ProDom; PD003639; DIX; 1.  
 DR ProDom; PD001580; Regl\_Gprotein; 1.  
 DR SMART; SM00315; RGS; 1.  
 DR PROSITE; PS50841; DIX; 1.  
 DR PROSITE; PS50132; RGS; 1.  
 KW Developmental protein; Wnt signaling pathway.  
 FT DOMAIN 54 172 RGS.  
 FT DOMAIN 640 646 POLY-SER.  
 FT DOMAIN 663 745 DIX.  
 FT CONFLICT 454 454 R -> Q (IN REF. 1).  
 FT CONFLICT 644 645 MISSING (IN REF. 1).  
 SQ SEQUENCE 745 AA; 81718 MW; 31A502528CEE84BA CRC64;

Query Match 88.1%; Score 37; DB 1; Length 745;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 655 QPPPLP 660

RESULT 15  
 DYN1\_CAEEL  
 ID DYN1\_CAEEL STANDARD; PRT; 830 AA.  
 AC P39055; Q93176; Q95QY9;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dynamin (EC 3.6.1.50).  
 GN DYN-1 OR C02C6.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=97439883; PubMed=9294229;  
 RA Clark S.G., Shurland D.L., Meyerowitz E.M., Bargmann C.I.,  
 RA van der Bliek A.M.;  
 RT "A dynamin GTPase mutation causes a rapid and reversible temperature-  
 RT inducible locomotion defect in *C. elegans*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:10438-10443(1997).  
 RN [2]  
 RP REVISIONS TO C-TERMINUS.  
 RA van der Bliek A.M.;  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
 RC STRAIN=Bristol N2;  
 RA Swinburne J.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Microtubule-associated force-producing protein involved  
 CC in producing microtubule bundles and able to bind and hydrolyze  
 CC GTP. Most probably involved in vesicular trafficking processes, in  
 CC particular endocytosis.  
 CC -!- CATALYTIC ACTIVITY: GTP + H(2)O = GDP + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Microtubule-associated.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=a;  
 CC IsoId=P39055-1; Sequence=Displayed;  
 CC Name=b;  
 CC IsoId=P39055-2; Sequence=VSP\_001329;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Belongs to the dynamin family.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L29031; AAB72228.2; -.  
 DR EMBL; Z79596; CAB01857.1; -.  
 DR EMBL; Z79596; CAC42251.1; -.  
 DR PIR; T18860; T18860.  
 DR HSSP; Q05193; 2DYN.  
 DR WormPep; C02C6.1a; CE07833.  
 DR WormPep; C02C6.1b; CE07832.  
 DR InterPro; IPR001401; Dynamin.  
 DR InterPro; IPR000375; Dynamin\_central.  
 DR InterPro; IPR003130; GED.  
 DR InterPro; IPR001849; PH.  
 DR Pfam; PF00350; dynamin; 1.  
 DR Pfam; PF01031; dynamin\_2; 1.  
 DR Pfam; PF02212; GED; 1.

DR Pfam; PF00169; PH; 1.  
 DR PRINTS; PR00195; DYNAMIN.  
 DR SMART; SM00053; DYNc; 1.  
 DR SMART; SM00302; GED; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PS00410; DYNAMIN; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KW Hydrolase; Motor protein; GTP-binding; Microtubule; Multigene family;  
 KW Endocytosis; Alternative splicing.  
 FT NP\_BIND 40 47 GTP (BY SIMILARITY).  
 FT NP\_BIND 138 142 GTP (BY SIMILARITY).  
 FT NP\_BIND 207 210 GTP (BY SIMILARITY).  
 FT DOMAIN 519 624 PH.  
 FT VARSPLIC 817 830 PGPGGPPPNMAPPR -> VPTPSNGAPEIPARPQVPKRPF  
 FT (in isoform b).  
 FT /FTId=VSP\_001329.  
 FT CONFLICT 734 734 R -> P (IN REF. 1).  
 SQ SEQUENCE 830 AA; 93407 MW; FF681250E51AB8A5 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QPPPLP 7  
 |||||  
 Db 747 QPPPLP 752

Search completed: August 24, 2004, 15:43:22  
 Job time : 9.76119 secs